Love and Money: A Theoretical and Empirical Analysis of Household Sorting and Inequality*

Raquel Fernández
NYU, CEPR and NBER

Nezih Guner
Queen's University

John Knowles
University of Pennsylvania

June 2001

Abstract

This paper examines the interactions between household matching, inequality, and per capita income. We develop a model in which agents decide whether to become skilled or unskilled, form households, consume and have children. We show that matches are increasingly correlated (sorted) in skill type as a function of the skill premium. In the absence of perfect capital markets, depending upon initial conditions, the economy can converge to steady states with a high degree of marital sorting, high inequality, and large fertility differentials or to ones with low sorting, low inequality and small fertility differentials. We use 34 country household surveys from the Luxembourg Income Study and the Inter-American Development Bank to construct several measures of the skill premium and of the degree of correlation of spouses' education (marital sorting). For all our measures, we find a positive and significant relationship between the two variables.

*We wish to thank Alberto Alesina, Oriana Bandiera, Jere Behrman, Alberto Bisin, Jason Cummins, Chris Flinn, Jonathan Portes, Gianluca Violante and David Weil for helpful comments and thank Miguel Szekely and Alejandro Gaviria at the IDB for assistance with the Latin-American data. For access to households surveys, we thank the LIS and the statistical agencies of the Latin American countries in our sample. The first and third authors acknowledge financial support from their respective NSF grants and the first author from the CV Starr Center as well.

1. Introduction

With a few notable exceptions, the analysis of household formation has played a relatively minor role in our understanding of macroeconomics. The vast majority of macroeconomic models tend to assume the existence of infinitely lived agents (with no offspring) or a dynastic formulation of mother (or father) and child (or children).[1] While this may be a useful simplification for understanding a large range of phenomena, it can also lead to the neglect of potentially important interactions between the family and the macroeconomy. This is especially likely to be the case in those areas in which intergenerational transmission plays a critical role, such as human capital accumulation, income distribution, and growth.

The objective of this paper is to examine some of the interactions between household matching ("marriage"), inequality (as measured by the skill premium), fertility differentials and per capita output. The main idea that we wish to explore, theoretically and empirically, is the potentially reinforcing relationship between the strength of assortative matching and the degree of inequality. In particular, we wish to examine the notion that a greater skill premium may tend to make matches between different classes (skilled and unskilled workers in our model) of individuals less likely, as the cost of "marrying down" increases. In an economy in which borrowing constraints can limit the ability of individuals to acquire optimal levels of education, this private decision may have important social consequences. In particular, it can lead to inefficiently low aggregate levels of human capital accumulation (resulting in larger wage inequality between skilled and unskilled workers), large fertility differentials across types of households, and lower per capita output. Thus, inequality and marital sorting are two endogenously determined variables that reinforce one another.

To explore the ideas sketched above, we develop a model in which individuals are either skilled or unskilled (according to education decisions made when young) and have a given number of opportunities in which to form a household with another agent. Once agents form households, they decide how much to consume and how many children to have. These children in turn decide whether to become skilled or unskilled workers. A decision to become skilled (synonymous here for acquiring a given level of education) is costly. To finance education, young individuals borrow in an imperfect capital market in which parental income plays the role of collateral. Thus parental income and the net return to being a skilled

[1] Even Becker and Tomes' (1979, 1986) pioneering work on intergenerational transmission of inequality assumes a one-parent household.

versus unskilled worker, including the expected utility from one's future match, determine the proportion of children that in aggregate become skilled. These individuals then also meet and form households, have children, and so on.

We show that the steady state to which this economy converges will in general depend upon initial conditions. In particular, it is possible to have steady states with a high degree of sorting (skilled agents form households predominantly with others who are skilled; unskilled form households predominantly with unskilled), high inequality, and large fertility differentials. Alternatively, there can be steady states with a low degree of sorting, low inequality and low fertility differentials.

Our empirical analysis examines the main implication of our model: a positive correlation between the skill premium and marital sorting. To do this, we assemble a total of 34 country household surveys from the Luxembourg Income Study (LIS) and the Inter-American Development Bank (IDB) and use them to construct a sample of households for each country. From these samples we construct several measures of the skill premium and two measures of the degree of correlation of spouses' education (our measure of marital sorting). For all our measures of the skill premium and marital sorting, we find a positive and significant relationship between the two variables.

Two other implications of our model are that greater marital sorting should imply lower per capita income and that the fraction of skilled labor in the economy and sorting should be negatively correlated across countries. We find evidence in favor of both of these predictions. Our model also implies that fertility differentials (between more and less educated households) should be increasing in inequality, a prediction that is borne out by the evidence presented by Kremer and Chen (1999).

Our work is related to several literatures. There is a rapidly growing literature on the intergenerational transmission of inequality in models with borrowing constraints. These models, though, either assume a dynastic formulation (e.g., Becker and Tomes (1986), Loury (1981), Ljungqvist (1993), Galor and Zeira (1993), Fernández and Rogerson (1998), Benabou (1996), Dahan and Tsiddon (1998), Durlauf (1995), Owen and Weil (1998) and Kremer and Chen (1999)) or consider a two-parent household in which the degree of sorting is exogenously specified (e.g. Kremer (1997) and Fernández and Rogerson (2001)). The last two papers are particularly relevant as they are concerned with whether an (exogenous) increase in marital sorting can lead to a quantitatively significant increase in inequality. In our model, on the other hand, sorting and inequality are endogenously determined. There is also a theoretical literature that focuses on the determinants of who matches with whom, but that basically abstracts from the endogeneity of the income distribution in the economy (e.g., Cole, Mailath, and Postlewaite (1992), and Burdett and Coles (1997, 1999)).[2] Our paper, therefore, is related to the two literatures, and can be seen as trying to integrate both concerns into a simple, analytical framework. Some recent work that also shares our concerns, but that is more focused on fertility, marriage and divorce, are Aiyagari, Greenwood, Guner (2000), Greenwood, Guner, and Knowles (2000), and Regalia and Rios-Rull (1999). The models, not surprisingly, are more complicated and rely on computation to obtain solutions for particular parameter values.

There is also a small, mostly descriptive, empirical literature that is related to our work. As reviewed by Lam (1988), the general finding in the literature is the existence of positive assortative matching across spouses. Mare (1991) documents the correlation between spouses' schooling in the US since 1930s. Using a large cross section of countries, Smith, Ultee, and Lammers (1998) find that the relation between marital sorting and some indicators for development (such as per capita energy consumption and the proportion of the labor force not in agriculture) has an inverted-U shape. Dahan and Gaviria (1999) report a positive relation between inequality and marital sorting for Latin American countries. Boulier and Rosenzweig (1984) document assortative matching with respect to schooling and sensitivity to marriage market variables using data from the Philippines.

2. The Model

In this section we present a model of matching, fertility and inequality. Each component of the model is kept relatively simple in the interest of highlighting the interactions among all three variables, both at a given moment in time and over the longer run.

2.1. Timing

The economy is populated by overlapping generations that live for two periods. At the beginning of the first period, young agents make their education decisions by deciding whether to become skilled or unskilled. This decision made, they then meet in what we call a "household matching market". Here they find another agent with whom to form a household, observing both the agent's skill type (and

[2] See Bergstrom (1997) and Weiss (1997) for a survey of the literature on theories of the family and household formation.

hence able to infer that agent's future income) and a match specific quality. They then enter into the labor market and work. In the second period, the agents, now adult, repay their education debt (if any), and households decide how much to consume and how many children to have.

We now describe in more detail each aspect of an agent's decision problem. We begin with the decision problem at the beginning of the second period, when agents have already formed a household of some given quality.

2.2. The Household's Problem

In this model we abstract from bargaining problems among agents within a household and instead assume that spouses share a common joint utility function.[3] We also abstract away from any differences between women and men, either exogenous (e.g., childbearing costs) or cultural/institutional (e.g., the degree of wage discrimination or the expected role of woman in the home relative to the workplace).[4]

Having matched in the first period of life and attained a match quality q, at the beginning of period 2 each household decides how much to consume, c, and how many children to have, n. Raising children is costly; each child consumes a fraction t of parental income, I.[5]

The utility of a household with match quality q and income I is given by solving:

$$\max_{c,n\geq 0}\left[c+\beta\log n+\gamma+q\right] \tag{2.1}$$

subject to

$$c\leq I(1-tn)$$

[3]For models that focus on intrafamily bargaining problems, see, for example, Bergstrom (1997) and Weiss (1997).

[4]This assumption considerably simplifies our analysis. See the conclusion for a brief discussion of alternative modelling assumptions.

[5]Traditionally the cost of having children is thought of as the opportunity cost of time. While in our model this interpretation is possible at the level of the individual budget constraint, we choose not to view it this way since, at the aggegate production function level, it is simpler if we do not have to take into account how hours of work vary across individuals (on account of different incomes implying different numbers of children). Instead we model the cost of children directly as a proportional consumption cost (perhaps as a result of bargaining in the household). An alternative route would have been to model a quality-quantity tradeoff in the production of children. We also allow the number of children to be a continuous rather than discrete variable to simplify the analysis.

where $\beta, t > 0$, and γ is a constant. Note that the way we have modelled match quality renders the solution to the optimization problem independent of q.

The household utility function implies that for household income below β, households will dedicate all their income to children and have $\overline{n} = \frac{1}{t}$ of them yielding utility $\beta \log \overline{n} + \gamma + q$. An interior solution to (2.1) is given by:

$$n = \frac{\beta}{tI} \tag{2.2}$$

and

$$c = I - \beta \tag{2.3}$$

Without loss of generality, by setting $\gamma = \beta \log t + \beta - \beta \log \beta$ we can write the indirect utility function for a couple with match quality q and household income $I > \beta$ as:

$$V(I, q) = I - \beta \log I + q, \text{ for } I > \beta \tag{2.4}$$

Note the comparative statics of the solution to the household's optimization problem. For values of household income below β, couples have a constant number of children and their utility is unaffected by increases in income within this range. For household income above β, increases in income increase consumption and reduce the number of children in the household. Thus, for $I > \beta$, wealthier households have fewer children and the fertility differential across income groups is increasing with income inequality.[6]

We next turn to the determination of household income.

2.3. The Labor Market

Agents are employed as workers in the second period of their lives. Workers are either skilled (s) or unskilled (u). We assume that technology is constant returns to scale and that wages are the outcome of a competitive labor market in which

[6] Fertility declining with income is consistent both with the cross-country evidence on fertility and per capita income (e.g. Perotti (1996)) and with cross-sectional evidence from US data (see Knowles (1998) and Fernández and Rogerson (forthcoming)). Furthermore, Kremer and Chen (1999) find that fertility differentials between low and high income families is increasing in the degree of inequality which is also implied by our model.

skilled and unskilled workers are employed to produce an aggregate consumption good.

Given a composition of the labor force L into skilled or unskilled workers ($L = L_s + L_u$), and denoting by λ the proportion of skilled workers in the population, full employment and constant returns to scale imply that output is given by:

$$F(L_s, L_u) \equiv LF(\lambda, 1-\lambda) \equiv L_u F(\frac{\lambda}{1-\lambda}, 1) \equiv L_u f(k)$$

where $k \equiv \frac{\lambda}{1-\lambda}$. Hence wages depend only on λ:

$$w_s = f'(k) \text{ and } w_u = f - f'k \tag{2.5}$$

We will often find it more convenient to work with the net return to being skilled which we denote by $\widetilde{w}_s \equiv w_s - d$, where d is the (constant) monetary cost of becoming skilled. Note that $\widetilde{w}_s$ is decreasing in λ, w_u is increasing in λ, and thus that the skill premium is a decreasing function of the fraction of skilled workers.

Household income I_{ij} is simply the sum of each partner's (i and j) wages. To simplify our analysis, we will assume that household income is always greater than β as this ensures an interior solution to the household maximization problem (as discussed in the previous section). We can do this either by imposing conditions on the production function such that the unskilled wage has a given positive lower bound of $\frac{\beta}{2}$ or by assuming that individuals are endowed with $e > \frac{\beta}{2}$ units of income. Thus, we assume:

$$2w_u > \beta. \tag{A1}$$

where w_u can be interpreted as the market wage (as in the first explanation) or as the market wage plus the endowment (as in the second explanation).

2.4. Household Matching

The choice of whom to match with is of course driven by many factors: tastes, one's environment (e.g., who one gets to know and the distribution of characteristics of individuals), and the prospects for one's material and emotional wellbeing. We provide a simple model in which we allow all these factors to interact to produce a household match.

Households can be categorized by the skill types of its two partners. Let I_{ij} denote the household income for a couple composed by skill types $i, j \in \{s, u\}$. Thus,

$$I_{ij} = \begin{cases} 2\widetilde{w}_s, & \text{if } ij = ss \\ \widetilde{w}_s + w_u, & \text{if } ij = su \\ 2w_u, & \text{if } ij = uu \end{cases} \tag{2.6}$$

We assume that in the first period, once their education decisions have been made, agents have two opportunities to match and form a household. In the first round, all agents meet randomly and draw a random match-specific quality q. This match can be accepted by both agents resulting in a "marriage" or rejected by at least one of the agents whereupon both agents enter the second round of matching. In the second round, agents are matched non-randomly with their own skill group and draw a new random match quality. We assume that qualities are match specific and are i.i.d draws from the same cumulative distribution function Q (with its pdf denoted by Q'), and with expected value μ and support $[0, \overline{q}]$.[7]

The two rounds of matching—one at random and the second exclusively with one's own skill type—are meant to reflect the fact that as time progresses one tends to meet people who are more like one in skill/education level (e.g., individuals who go on to college meet other people also in college, whereas individuals who work in low-skill jobs tend to have more contact with other individuals of the same skill level).[8] Note that a skilled agent (with a high wage) that encounters an unskilled agent (with a low wage) in the first round and draws a high q will face a tradeoff between forming a lower-income household with a high quality match and a higher-income household (by matching for sure with a skilled agent in the second round) but of an unknown quality (i.e., there is a tradeoff of "love versus money").

[7] The assumption of $q \geq 0$ ensures that all agents will form a household in the second round. Although unrealistic, this allows us to abstract from the issue of how inequality affects the decision to remain single, which is not the focus of the analysis here. In our comparative static analysis, we will assume that $\overline{q}$ is sufficiently large so that in equilibrium some matches occur between skilled and unskilled individuals. This is for simplicity only.

[8] Alternative modelling assumptions (e.g., more periods, search or waiting costs, and assuming individuals always meet others at random) are also possible and can give rise to similar properties as this one. This formulation is simple and avoids problems of multiple equilibria that can arise when the fraction of types an individual meets evolves endogenously over time. See Fernández and Pissarides (2001) for an infinite horizon search model for household partners.

Let $V_{ij}(q)$ denote the utility of a couple with income I_{ij} and match quality q (as expressed in (2.4)) where $i, j \in \{s, u\}$. As a skilled agent's second-round option dominates that of an unskilled agent (given $\widetilde{w}_s \geq w_u$, which is a necessary condition in order for any individual to choose to become a skilled worker), it is the skilled agent who determines whether a match between a skilled and an unskilled agent is accepted.

A skilled agent is indifferent between accepting a first-round match with an unskilled agent and proceeding to the second round if $V_{su}(q) = V_{ss}(\mu)$. Solving for the level of q at which this occurs, q^*, yields a threshold quality of:

$$q^* = I_{ss} - I_{su} - \beta \log\left(\frac{I_{ss}}{I_{su}}\right) + \mu \tag{2.7}$$

which, after substituting for wages, yields:

$$q^*(\lambda) = \widetilde{w}_s(\lambda) - w_u(\lambda) - \beta \log\left(\frac{2\widetilde{w}_s(\lambda)}{\widetilde{w}_s(\lambda) + w_u(\lambda)}\right) + \mu \tag{2.8}$$

The intuition underlying (2.8) is clear. A skilled individual who matches with an unskilled one in the first round knows that by foregoing that match she will meet a skilled individual in the second round with an expected match quality of μ. Thus, the match quality of the unskilled individual must exceed μ by the amount required to compensate for the decreased utility arising from the fall in household income. Of course, the threshold quality for two agents of the same type to match in the first round is μ as this is the expected value of next round's match quality and there is no difference in household income.

Given a distribution of individuals into skilled and unskilled, we can find the fraction of households that will be composed of two skilled individuals, two unskilled individuals, and one skilled and one unskilled. The fraction of households of each type depends only on the probability of types meeting in the first round and on q^*. Both of these are only a function of λ_t since this variable determines both household incomes and first round matching probabilities. Denoting by ρ_{ij} the fraction of households formed between agents of skill type i and j, $i, j \in \{s, u\}$ (with $\rho_{su} = \rho_{us}$), these are given by:

$$\rho_{ij}(\lambda_t) = \begin{cases} \lambda_t^2 + \lambda_t(1-\lambda_t)Q(q^*(\lambda_t)), & \text{if } ij = ss \\ 2\lambda_t(1-\lambda_t)(1-Q(q^*(\lambda_t))), & \text{if } ij = su \\ \lambda_t(1-\lambda_t)Q(q^*(\lambda_t)) + (1-\lambda_t)^2, & \text{if } ij = uu \end{cases} \tag{2.9}$$

How does a change in proportion of skilled workers in the population affect the fraction of households of each type? An increase in the λ will unambiguously decrease the fraction of couples that are uu as, for any given q^* they are less likely to end up in uu households. Furthermore, q^* will decrease, thereby increasing the probability that a first round match between a high and low skilled worker will result in a household. The effect on us and ss households, on the other hand, is ambiguous (although the aggregate fraction of the population that is in one of these two types of households must of course increase). For any given q^*, the fraction of ss households increases, but as a skilled individual is now more willing to match with an unskilled one, this will work to decrease the fraction of ss households. The effect on us households is positive if $\lambda \leq 1/2$ (as both the likelihood of s and u individuals meeting in the first round increases as does the probability that the match will be accepted) and ambiguous otherwise.

Note that $Q(q^*)$ is a measure of the degree of sorting that occurs. If individuals were not picky and simply matched with whomever they met in the first round, then q^* would equal zero and ρ_{su} would equal the probability of a skilled and an unskilled individual meeting, i.e., $2\lambda_t(1-\lambda_t)$. If individuals simply cared about quality and not about income, then q^* would equal μ. Lastly, if individuals cared only about income and not about match quality, then $Q(q^*)$ would equal one and there would be no matches between skilled and unskilled agents.

Remark 1. *$Q(q^*)$ is the correlation coefficient between different skill types in households.*

The observation above will be very useful when we examine the data as although the fractions of couples of each type that form may have ambiguous comparative statics with respect to λ, this is not true for the degree of sorting (i.e., for the correlation coefficient). This is stated in the theorem below.

Theorem 2.1. *An increase in λ will decrease the degree of sorting.*

Proof: Recall that the degree of sorting is given by $Q(q^*)$. Note that

$$\frac{\partial q^*}{\partial \widetilde{w}_s} > 0, \qquad \frac{\partial q^*}{\partial w_u} < 0 \tag{2.10}$$

and $\frac{d\widetilde{w}_s}{d\lambda} = f''\frac{dk}{d\lambda} < 0$, $\frac{d\widetilde{w}_u}{d\lambda} = -f''k\frac{dk}{d\lambda} > 0$, where $\frac{dk}{d\lambda} = \frac{1}{(1-\lambda)^2}$. Hence, $\frac{dQ(q^*)}{d\lambda} < 0$. ||

Note that the theorem above also implies (by (2.10)) that an exogenous increase in inequality (say, from a technology shock) will also increase sorting by making skilled workers less willing to form households with unskilled workers.

The main focus of our empirical work will be in establishing the positive correlation implied above between the skill premium and the degree of sorting in households across countries. Why should different countries have different degrees of inequality, however? This is the question that the model next turns to by examining the determinants of a young agent's decision to become a skilled relative to an unskilled worker.

2.5. Education Decisions and Capital Markets

A young agent's desire to become skilled depends on the return to being a skilled relative to an unskilled worker. Note that this depends not only on net wages next period, but also on the expected return to matching at the household level. The expected value of being a skilled worker given that a fraction λ_{t+1} of the population also becomes skilled is given by:

$$\begin{aligned} V^s(\lambda_{t+1}) &= \lambda_{t+1} \int_0^{\overline{q}} \max\left[V_{ss}(x;\lambda_{t+1}), V_{ss}(\mu;\lambda_{t+1})\right] dQ(x) \\ &\quad +(1-\lambda_{t+1}) \int_0^{\overline{q}} \max\left[V_{su}(x;\lambda_{t+1}), V_{ss}(\mu;\lambda_{t+1})\right] dQ(x) \end{aligned}$$

whereas the expected value of being an unskilled worker is:

$$\begin{aligned} V^u(\lambda_{t+1}) &= \lambda_{t+1}[\int_0^{q*} V_{uu}(\mu;\lambda_{t+1}) dQ(x) + \int_{q*}^{\overline{q}} V_{su}(x;\lambda_{t+1}) dQ(x)] \\ &\quad +(1-\lambda_{t+1}) \int_0^{\overline{q}} \max\left[V_{uu}(x;\lambda_{t+1}), V_{uu}(\mu;\lambda_{t+1})\right] dQ(x) \end{aligned} \tag{2.11}$$

We assume that in addition to a monetary cost of d, becoming a skilled worker entails an additive non-pecuniary cost of $\delta \in [0, \infty]$. This cost is assumed to be identically and independently distributed across all young agents with cumulative distribution function Φ. Thus, an agent with idiosyncratic cost δ_i will desire to become skilled if $V^s - V^u \geq \delta_i$.

We define by $\delta^*(\lambda)$ the skilled-unskilled payoff difference generated when a fraction λ of the population is skilled, i.e.,

$$\delta^*(\lambda_{t+1}) \equiv V^s(\lambda_{t+1}) - V^u(\lambda_{t+1}) \tag{2.12}$$

Note that given δ^*, all agents with $\delta_i \leq \delta^*$ would want to become skilled. If young agents were able to borrow freely, children from all household types would make identical education decisions contingent only on their value of δ_i. Hence in equilibrium a fraction $\Phi(\ \delta^*)$ of each family would become skilled yielding $\lambda_{t+1} = \Phi(\ \delta^*)$ and

$$\delta^*(\Phi(\delta^*)) \equiv V^s(\Phi(\delta^*)) - V^u(\Phi(\delta^*)) \tag{2.13}$$

If, however, parental income is a factor that influences a child's access to capital markets (either in terms of the interest rate faced or in determining whether they are rationed in the amount they are able to borrow), then children of different household types may make different education decisions although they have the same δ_i. In this case, the fraction of children of different household types that become skilled will depend on the parental household income distribution, and thus on λ_t.[9]

In particular, we assume that children within a family with household income I can borrow on aggregate up to $Z(I)$, $Z' > 0$. One way to think about this constraint is that parents can act as monitoring devices for their children in an incentive compatible fashion by putting their own income up for collateral. This ensures that the children will use the funds to become educated rather than for consumption and allows up to $Z(I)$ to be borrowed by the family's children. Hence, a family with income I and $n(I)$ children can at most afford to educate at a cost d per child a fraction $\Phi(\widehat{\delta}(I))$ implicitly defined by:[10]

$$\frac{Z(I)}{n(I)\Phi(\widehat{\delta})} = d \tag{2.14}$$

Note that as indicated in (2.14), children from families with low household income are hampered in their ability to become skilled both because of the lower aggregate

[9] It is important to note that this constraint should not be interpreted literally as the inability to borrow freely to attend college. Instead, it is best thought of as a shorthand for parental inability to borrow against their children's future human capital so as to live in a neighborhood in which the quality of primary and secondary public education is high or to opt out of public education for a high-quality private education. It is the quality of this earlier education that then determines the probability of an individual attending college even if the latter is free.

[10] We are implicitly normalizing the gross interest rate to equal one. Note that as we are not endogenizing the supply of funds for loans, it is best to think of loans being provided on a world market (in which this country is small).

amount that can be borrowed by the family and because of the larger number of children (recall that n is decreasing in I) that want to become skilled and hence must share these funds.

Thus, given λ_t (and hence family income and number of children by family type), in equilibrium a fraction

$$\pi_{ij}(\lambda_t, \lambda_{t+1}) \equiv \min[\Phi(\delta^*(\lambda_{t+1})), \Phi(\widehat{\delta}(I_{ij}(\lambda_t)))] \tag{2.15}$$

of each family type will become skilled.[11]

2.6. Equilibrium

Given a division of the young population into skilled and unskilled in period t, i.e., λ_t , an equilibrium for that period is a skilled and unskilled wage pair $(w_s(\lambda_t), w_u(\lambda_t))$ given by (2.5), a threshold match quality (between skilled and unskilled agents) of $q^*(\lambda_t)$ given by (2.8), which generates a division of families into types $\rho_{ij}(\lambda_t)$ as given by (2.9). It also includes a decision by the children of these individuals to become skilled or unskilled next period such that given that the expected value of λ in the next period is λ_{t+1} and hence the differential between the expected value of being a skilled or unskilled worker is $\delta^*(\lambda_{t+1})$, a fraction $\pi_{ij}(\lambda_t, \lambda_{t+1})$ given by (2.15) of each family type becomes skilled, and in aggregate these constitute a proportion λ_{t+1} of next period's population.

Figure 1 depicts the equilibrium λ_{t+1} generated by a given λ_t. The upward sloping line, $\delta = \Psi(\lambda_{t+1}; \lambda_t)$, is derived in the following fashion. For a given λ_t, it shows what δ would have to be such that the fraction of young individuals with $\delta_i \leq \delta$ that would be able to afford to enter the following period as skilled equals λ_{t+1}. Note that the domain of this function can in general be smaller than 1 since for some initial conditions not all individuals will be able to afford to become skilled even if $\delta \to \infty$. In the absence of borrowing constraints, the inverse of this function would coincide with $\Phi(\delta)$ and the unconstrained $\Psi(\cdot)$ curve is the lower envelope of the family of curves parameterized by different values of λ_t. The downward sloping curve shows $\delta^*(\lambda_{t+1}) \equiv V^s(\lambda_{t+1}) - V^u(\lambda_{t+1})$ as a function of λ_{t+1}. Note that this curve does not depend on λ_t. The intersection of these two curves gives the equilibrium values of $(\delta^{**}, \lambda_{t+1}^{**})$ given λ_t.

Existence of an interior equilibrium (for any initial λ_t) is guaranteed if we assume that $\widetilde{w}_s(\lambda) < w_u(\lambda)$ for some $\lambda \in (0, 1)$ (i.e., such that no one finds it in

[11] We are assuming that the decision regarding which children should obtain the funding to become skilled is efficient, i.e., those who have the lowest δ are the first to become skilled.

their interest to become skilled) and that for some other $\lambda \in (0,1)$ the inequality is reversed. Note that the Ψ curve is continuous, upward sloping, starts at zero, and becomes vertical oncle all family groups are constrained. Thus, this and the fact the $\delta^*(\lambda_{t+1})$ is a continuous function defined over the entire range of $[0,1]$ and goes from strictly positive to strictly negative numbers, guarantees the existence of an interior equilibrium. Uniqueness of equilibrium (for any given λ_t) is guaranteed if δ^* is monotonically decreasing in λ.

2.7. Inequality

In order to investigate the effects of inequality on household sorting and education decisions, we first examine how exogenous changes in inequality affect education choices in any given period (i.e., we examine the effect of changes in wages taking λ as given).

An increase in $\widetilde{w}_s$ makes becoming a skilled worker more attractive as it increases the direct return to being skilled. It also increases the return to matching with another skilled worker, making skilled agents pickier in their household matching, i.e., it increases q^*. On the other hand, an increase in $\widetilde{w}_s$ has ambiguous effects on an unskilled agent's payoff since although it increases the value of being in a household with a skilled worker, it also makes these matches more unlikely. It is easy to show that an increase in $\widetilde{w}_s$ increases the relative desirability of being a skilled relative to an unskilled worker, i.e.,

$$\begin{aligned}\frac{d\delta^*}{d\widetilde{w}_s} &= \frac{d[V^s - V^u]}{d\widetilde{w}_s} \\ &= [\lambda + (1-\lambda)Q(q^*)]\frac{\partial V_{ss}}{\partial \widetilde{w}_s} + [(1-Q(q^*))(1-2\lambda)]\frac{\partial V_{su}}{\partial \widetilde{w}_s} \\ &\quad -\frac{dq^*}{d\widetilde{w}_s}\lambda[V_{uu}(\mu) - V_{su}(q^*)]Q'(q^*)\end{aligned}$$

which is strictly positive as $\frac{\partial V_{ss}}{\partial w_s} = 2 - \frac{\beta}{w_s} > \frac{\partial V_{su}}{\partial w_s} = 1 - \frac{\beta}{w_s + w_u}$, $\lambda + (1-\lambda)Q(q^*) > (1-Q(q^*))(1-2\lambda)$, $\frac{dq^*}{dw_s} > 0$ and $V_{uu}(\mu) - V_{su}(q^*) < 0$ (with the latter following from the fact that skilled workers choose a higher cutoff quality level in their matches with unskilled individuals than what the latter find optimal).[12]

An increase in w_u, on the other hand, has ambiguous effects on the relative desirability of being a skilled worker relative to an unskilled worker, as

[12]For notational convenience, we have supressed everywhere the dependence of V_{ij} on λ.

$$\begin{aligned}\frac{d\delta^*}{dw_u} &= [(1-Q(q^*))(1-2\lambda)]\frac{\partial V_{su}}{\partial w_u} - [1-\lambda+\lambda Q(q^*)]\frac{\partial V_{uu}}{\partial w_u} \qquad (2.16)\\ &+\frac{dq^*}{dw_u}\lambda[V_{su}(q^*)-V_{uu}(\mu)]Q'(q^*),\end{aligned}$$

The expression on the second line is negative but the expression on the first line, which can be written as $-1-Q(q^*)+\frac{\beta}{w_u(w_u+w_s)}[w_s(\lambda Q(q^*)+1-\lambda)+w_u(Q(q^*)+\lambda(1-Q(q^*))]$ is ambiguous.[13]

In our model, of course, wages do not change exogenously but instead respond to changes in λ. We next turn to an analysis of the effect of an increase in λ on the relative attractiveness of becoming skilled. Note that a change in the fraction of the population that plans to become skilled will have two effects (i) it will change wages and hence household incomes by changing the ratio of skilled to unskilled workers in aggregate production; (ii) it will change the probability with which individuals encounter skilled relative to unskilled workers in the first round of matching, (i.e., λ). So, the total effect on the payoff differential δ^* between skilled and unskilled agents is given by:

$$\frac{d\delta^*}{d\lambda}=\frac{\partial[V^s-V^u]}{\partial\widetilde{w}_s}\frac{d\widetilde{w}_s}{d\lambda}+\frac{\partial[V^s-V^u]}{\partial w_u}\frac{dw_u}{d\lambda}+\frac{\partial\delta^*(\lambda)}{\partial\lambda}$$

Note that we can rewrite δ^* as:

$$\begin{aligned}\delta^* &= \lambda\left[V_{ss}(\mu)Q(\mu)+\int_{\mu}^{\overline{q}}V_{ss}(x)dQ(x)-\int_{q^*}^{\overline{q}}V_{su}(x)dQ(x)-V_{uu}(\mu)Q(q^*)\right]\\ &+(1-\lambda)\left[V_{ss}(\mu)Q(q^*)+\int_{q^*}^{\overline{q}}V_{su}(x)dQ(x)-V_{uu}(\mu)Q(\mu)-\int_{\mu}^{\overline{q}}V_{uu}(x)dQ(x)\right].\end{aligned}$$

which after substituting in (2.4) and (2.6) yields:

$$\begin{aligned}\delta^* &= (\widetilde{w}_s-w_u)(1+Q(q^*))+(2\lambda-1)\int_{\mu}^{q^*}(x-\mu)dQ(x)\\ &+(2\lambda-1)(1-Q(q^*))\beta\log(\widetilde{w}_s+w_u)\\ &-(\lambda+(1-\lambda)Q(q^*))\beta\log 2\widetilde{w}_s+(1-\lambda+\lambda Q(q^*))\beta\log 2w_u.\end{aligned}$$

[13]This ambiguity is due to the fact that an increase in w_u also makes a skilled worker better off (as the return to matching with an unskilled individual increase) and, as our indirect utility function in convex in income, this effect could in theory outswamp the direct effect of the increase in w_u on V^u.

Taking the derivative of δ^* with respect to λ yields (after some manipulation):

$$\begin{aligned}\frac{d\delta^*}{d\lambda} &= R\{(1+2kQ(q^*)+k^2)(\widetilde{w}_s+w_u-\beta)+(2k+Q(q^*)k^2+Q(q^*))(\widetilde{w}_s+w_u) \\ &\quad -(k+Q(q^*))\beta\frac{w_u}{\widetilde{w}_s}-(k+k^2Q(q^*))\beta\frac{\widetilde{w}_s}{w_u}\}+ \qquad (2.17) \\ &\quad Q'(q^*)\frac{\partial q^*}{\partial\lambda}\lambda\left[2(\widetilde{w}_s-w_u)-\beta\log\frac{\widetilde{w}_s}{w_u}\right]+\{2\int_{\mu}^{q^*}(x-\mu)dQ(x) \\ &\quad +\beta\left[2\log(\widetilde{w}_s+w_u)-\log(2\widetilde{w}_s)-\log(2w_u)\right]\left[1-Q(q^*)\right]\}.\end{aligned}$$

where $R=\frac{f''}{(\widetilde{w}_s+w_u)}\frac{1}{(1+k)}\frac{dk}{d\lambda}<0$.

In order to sign $\frac{d\delta^*}{d\lambda}$, note that all expressions other than the last one in curly brackets are negative. To see this, note that, as shown in Appendix A, the sign of the expression in the first curly parenthesis (the first two lines) of (2.17) is positive (which, as multiplied by $R<0$ implies that the first two lines are negative) and that $\frac{\partial q^*}{\partial\lambda}<0$ (and the expression multiplying it is positive). Unfortunately, we are not unambiguously able to sign the equation as the effect of a change in λ on the matching component is strictly positive (i.e., $\frac{d\lambda}{dk}>0$, $\int_{\mu}^{q^*}(x-\mu)dQ(x)>0$, and the expression on the fourth line is positive since $\log x$ is a concave function).

The ambiguity in (2.17) above is due to the fact that although an increase in λ decreases skilled wages and increases unskilled wages, thereby making it less attractive to become skilled than previously, it also increases the probability of matching with a skilled agent in the first round. As the indirect utility function is convex in income, then for a given cutoff level of q^*, the increased probability of meeting a skilled individual on the margin yields greater utility to another skilled individual.

In what follows, in order to ensure the existence of a unique equilibrium, we will assume that:[14]

$$\frac{d\delta^*(\lambda)}{d\lambda}<0 \qquad \text{(A2)}$$

2.8. Steady States and Dynamics

The state variable for this economy is the fraction of skilled workers, λ. The evolution of this variable is given by:

[14] We simulated the model for various functional forms and parameter values. We always found equilibrium to be unique.

$$\lambda_{t+1}(\lambda_t, E\lambda_{t+1}) = \frac{L_{s,t+1}(\lambda_t, E\lambda_{t+1})}{L_{t+1}(\lambda_t)} \tag{2.18}$$

We discuss each component of this equation in turn.

The population at time $t+1$ is simply the sum over all the children born to households in period t. Hence,

$$L_{t+1}(\lambda_t) = [n_{ss}(\lambda_t)\rho_{ss}(\lambda_t) + n_{su}(\lambda_t)\rho_{su}(\lambda_t) + n_{uu}(\lambda_t)\rho_{uu}(\lambda_t)]L_t \tag{2.19}$$

where $n_{ij}(\lambda)$ is the utility maximizing number of children for a household with income $I_{ij}(\lambda)$ as indicated in equation (2.2).

The skilled population at time $t+1$ is simply the sum over all children born to households in period t who decide to become skilled. Recall that some household types may be constrained and hence that the decision to become skilled depends (potentially) both on parental income in period t and hence on λ_t as well as on payoffs expected for $t+1$ (and hence on $E_t\lambda_{t+1}$, where E is the expectations operator).[15] Thus,

$$\begin{aligned} L_{s,t+1}(\lambda_t, \lambda_{t+1}) &= [\pi_{ss}(\lambda_t, \lambda_{t+1})\, n_{ss}(\lambda_t)\rho_{ss}(\lambda_t) + \pi_{su}(\lambda_t, \lambda_{t+1})\, n_{su}(\lambda_t)\rho_{su}(\lambda_t) \\ &\quad + \pi_{uu}(\lambda_t, \lambda_{t+1})\, n_{uu}(\lambda_t)\rho_{uu}(\lambda_t)]L_t \end{aligned} \tag{2.20}$$

A steady state is defined as a $\lambda_t = \lambda^*$ such that $\lambda_{t+1}(\lambda^*, \lambda_{t+1}) = \lambda^*$. Note that if λ is constant, so are wages, and so is the cutoff quality for a skilled agent to match with an unskilled agent and the education decisions of children.

If the economy had perfect capital markets, then independently of the initial value of λ, the ability of individuals to borrow implies that a fraction $\widetilde{\lambda} = \Phi(\widetilde{\delta})$ of them will choose to become skilled, i.e. $\pi_{ij} = \widetilde{\lambda}$, $\forall ij, \forall \lambda_t$ such that $\delta^*\left(\widetilde{\lambda}\right) = \widetilde{\delta}$. Thus the economy would converge immediately to the unique steady state.

In the absence of perfect capital markets, the initial distribution of individuals into skilled and unskilled determines the dynamic evolution of the economy. With borrowing constraints, for those family types who are constrained, a fraction

[15] Rational expectations implies that in equilibrium $E_t\lambda_{t+1} = \lambda_{t+1}$, so we have suppressed the expectations operator in what follows.

smaller than $\Phi(\delta^*)$ will be able to become skilled, and thus in aggregate a fraction that is smaller than $\Phi(\delta^*)$ will become skilled next period. Obviously, the first family type to be constrained will be the uu type, followed by the us type and lastly by the ss type, as lower family income implies both more binding borrowing constraints and a larger number of children who wish to borrow.

As shown in Figure 2 for a particular CES production function, this economy can easily give rise to multiple steady states, here given by all the intersections of λ_{t+1} with the 45 degree line.[16] As depicted in the figure, the steady states A and B are locally stable.[17] The steady state in A is characterized by a low fraction of skilled individuals, high inequality between skilled and unskilled workers, much sorting in household formation (i.e., skilled individuals predominantly marry other skilled ones; unskilled individuals predominantly marry other unskilled), and high fertility differentials (i.e., $\frac{n_{uu}}{n_{ss}} = \frac{I_{ss}}{I_{uu}}$ is high). In the steady state B, the opposite is the case: there is a large fraction of skilled individuals, low inequality, low sorting and low fertility differentials.

Across steady states and indeed across any equilibrium at a point in time, higher inequality is associated with higher sorting. This follows simply from the static analysis in which we showed that greater wage differentials imply greater sorting (Theorem 2.1). What we would also like to be able to show is that (out of steady state) economies that start out with greater inequality end up in a steady state with at least as much inequality, sorting, and fertility differentials than an economy that starts out with lower inequality. This we have confirmed for a large number of simulations but have so far been unable to prove analytically. This does not affect, however, the prediction which we will examine in the data: the existence of a positive correlation between sorting and the skill premium. We now turn to our empirical analysis.

[16] The functional forms used to generate this figure are a production function given by $F(L_s, L_u) = (\alpha L_s^\gamma + (1-\alpha)L_u^\gamma)^{1/\gamma}$, and a limit on aggregate borrowing by children within a family of a fraction θ of household income, i.e., $Z(I) = \theta I$. Lastly, we assume that δ is distributed uniformly and that q is distributed with a triangular density function. The parameter values used are: $\alpha = 0.2$, $\gamma = 0.5$, $\theta = 0.1$ $\overline{\delta} = 0.2$, $\overline{q} = 8$, $\beta = 0.05$, $t = 0.05$, and $d = 0.1$.

[17] Note that the number of locally stable steady states can be greater than two since this depends on the change in the fraction of children of different families types that are constrained at different values of λ.

3. Empirical Analysis

The basic prediction of our model is the existence of a positive relation between the skill premium and the degree of household sorting. This relationship should hold independently of whether countries have the same technology or whether they are converging to the same or different steady states. The purpose of this section is to establish that there is indeed a positive correlation between marital sorting and the skill premium across different countries, and that this correlation is robust with regards to the main concerns that arise with regards to the data. Our data set does not allow us to examine causality since we are not able to identify exogenous variations in either of these two variables. Hence, our basic results regarding the correlation between marital sorting and the skill premium are based on OLS regressions of the skill premium on marital sorting (although obviously these regressions can be run the other way around as well).[18]

We examine the main implication of our model using household surveys from 34 countries in various regions of the world.[19] For each country we assemble a sample of households with measures of the education and wages of both spouses. We then construct several measures of the skill premium for high-skill workers and two measures of the degree of marital sorting by education for each country. We use these measures to examine the correlation between the skill premium and sorting across countries.

We find a positive and significant relation between the skill premium and marital sorting, and show that this finding is robust to the partitioning of the sample into a subsample for Latin America and one for the rest of the world. If countries have the same technology, our model also predicts that countries with a high degree of sorting should also have a relatively low level of GDP per capita. Furthermore, countries with a high degree of sorting, ceteris paribus, should have a low fraction of skilled workers. We find evidence in favor of both of these negative relationships. Altogether we take these findings to suggest agreement of our basic hypotheses with the data.

3.1. Sample

The data consists of a collection of household surveys assembled from the Luxembourg Income Study (LIS) and a collection of Latin-American household surveys

[18] All of our basic results hold when we run regressions of sorting on the skill premium instead.

[19] A larger sample would be desirable, but there are few countries for which these household data sets are available.

held by the Inter-America Development Bank (IDB). From the LIS we obtain wage and education data at the household level for 20 countries, largely European, but also including Australia, Canada, Israel, Taiwan and the U.S.[20] The years of these surveys ranges from 1990 to 1995. For Britain we use the British Household Panel Study (1997), rather than the data from the LIS, since in the latter the education variable is reported as the age at the completion of education, a variable that was hard to map into years of schooling. The 13 IDB countries are all located in Latin America and the surveys date from 1996-1997. We provide a more detailed discussion of these household surveys in Appendix B, where we list the names and sample sizes of the surveys by country.

For each country we construct a sample of couples where the husband is between 36 to 45 years old.[21] We include households in the analysis if, in addition to the age requirement, there is a spouse present and education and earnings variables are available for both spouses. To avoid problems of income attribution across multiple families within a household, the sample is further restricted to couples where the husband is the head of the household in the Latin-American countries, and to single-family households in the LIS surveys.[22] We do not restrict the definition of a spouse to legally married couples, but for convenience we refer to them as "wives" and "husbands". In addition to this main sample, we also use the sample of husbands of ages 30-60 years to check the robustness of our skill-premium results.

We use labor income as our measure of the return to education. All of the surveys report income of each spouse, though the details of what is reported differs by country. Some LIS countries report gross annual labor earnings, all forms of cash wage and salary income, and some report these net of taxes. Income in the Latin American countries is gross monthly labor income from all sources. This definition includes income from both primary and secondary labor activities; the

[20]Russia is also available in the LIS, but we choose not to include it due to the low quality of the data. Our basic results in any case hold if Russia is included.

[21]The analysis was restricted to a narrow cohort for a few reasons. First it makes the issue of how to control for age less important in estimating the effect of education on earnings, so that simple measures like ratios of averages do not reflect noise from demographic variation. Second, as wage premia may change over time, the observed wage premium is presumably a better measure of younger workers' perception at the time of "marriage" decisions. Finally, for workers who are much younger, the observed wage is less closely related to lifetime earnings. For some of our inequality measures, however, we look at all couples in the surveys who are between the ages of 30 and 60.

[22]We were not able to reliably identify all multi-household families in the Latin American surveys, so we cannot explicitly eliminate multi-family households in these countries.

exact components vary somewhat across countries, but generally include wages, income from self-employment, and proprietor's income, as well as adjustment to reflect imputation of non-monetary income. Appendix B provides the details of our income measures. The fact that some countries report gross income while others report net income could distort our cross-country comparisons, as net income will be more equally distributed than gross income in those countries with progressive taxation. To attempt to deal with this we introduced a dummy variable in our basic regressions representing whether net or gross income is reported in a country. This variable did not affect our basic results, indicating that they were not driven by this particular feature of the data.

Like income, education measures also differ across countries. While education in the Latin American data is reported as total years of schooling, and in some cases the highest level attained, in the LIS countries the education units are quite idiosyncratic. Some countries report years, while others report attainment by country-specific levels. We attempt to standardize the LIS education variable by converting the reported units to years of education. In addition, we create a skill indicator variable that equals 1 if an individual has years of schooling that exceed high-school completion level and equals zero otherwise. This requires us to determine how many years of schooling an individual needs to be able to go beyond high school in each country. The Latin American data also required some standardization because the number of years required for high-school completion varies across countries. For countries that report attainment together with years of schooling, our skill indicator equals 1 if some post-secondary education was reported for an individual. For countries that do not report attainment level, our skill-indicator equals 1 if the years of schooling exceeded the standard time required to complete high school in that country. Our mapping of reported education measures into years of schooling and into an indicator for high school completion are summarized in Table B2 of Appendix B.

3.2. Variables

We construct two basic measures of the skill (education) premium for each country. The first is the ratio of earnings for skilled male workers to unskilled ones in our sample, i.e. husbands between ages 36 and 45.[23] This measure is very simple and

[23] We focus primarily on the male skill premium as women's labor supply decision is more likely to depend on her spouse's earnings. See the conclusion for a discussion of how our model can be modified to deal with this consideration.

intuitive, and has a direct counterpart in our model. A potential drawback of using the wage ratio as described above is that it reflects income at a particular stage in the life-cycle, and the mapping from this variable to lifetime income is likely to differ across skill groups. It also ignores information other than education that could also affect earnings, such as age or labor market experience. We control for such effects by constructing another measure of the skill premium; this is the coefficient on an indicator for being skilled (i.e., having at least some post high-school education) in the following regression:

$$\log(e_i) = a_0 + a_1 I_i + a_2(age - s_i - 6) + a_3(age - s_i - 6)^2 + \varepsilon_i,$$

where e_i is the earnings, I_i is an indicator for being skilled, s_i is years of schooling, and $(age - s_i - 6)$ is potential experience for individual-i. This regression is estimated for each country by OLS for all husbands aged 30-60 who have positive earnings rather than solely for those aged 36-45. Given that we have controlled for experience, this measure may be able to better capture potential lifetime labor earnings inequality than the simple ratio of earnings for our smaller sample.[24] We will refer to this measure as the skill indicator measure of inequality and to the previous one as the wage ratio measure of inequality. To summarize, these two measures will differ as the skill indicator uses a larger sample, omits zero-earnings and controls for experience.

Our main measure of sorting is the Pearson correlation coefficient between husband's and wife's years of education across couples in our sample. We call this the sample correlation measure of marital sorting. An alternative measure is given by the rank correlation between years of schooling of spouses which we use to check the robustness of our results.

Table 1 reports the measures of the skill premium and sorting for each country. The first column reports the means and standard deviation of the fraction of skilled husbands in our sample for each country. The column labelled "Skilled Share" gives the percent of the sample with more than high-school education. The mean level of the share of skilled husbands across countries in our sample is 23.7% with a standard deviation of around 12.6%. The second and third columns show the wage ratio and skill indicator measures of the skill premium. The average level of the wage ratio across countries is about 2 with a standard deviation of around 0.88; the same statistics for the skill-indicator measure are 0.48 and 0.21, respectively.

[24] How good this measure is of lifetime labor earnings inequality depends on how well the earnings of different cohorts at a point in time represents the lifecycle earnings of an individual (i.e., on the stability of the earnings profile).

The last two columns report the sample and rank correlation measure of marital sorting. On average, the sample correlation between spouses' years of schooling is about 0.59 with a standard deviation of 0.12. The countries with the lowest skill premia are Australia and Denmark (wage ratio) and Israel (skill indicator measure), while Colombia and Brazil (wage ratio) and Chile and Brazil (skill indicator measure) have the highest. The correlation of the years of schooling across spouses is lowest for Australia, and highest for Colombia and Ecuador.

Table 2 shows the correlation among all our variables. Our two main measures of the skill premium are highly correlated (0.84), as are our two measures of marital sorting (0.96) and the measures of marital sorting and the skill premium (around 0.6 in each case). All of the correlations are significant at the 1% level.

3.3. Results

This section reports the main results of our empirical analysis. Table 3 shows the results from a regression of marital sorting on the skill premium. In Table 3(a), the dependent variable is the wage ratio, and the explanatory variable is the sample correlation between husband's and wife's education. The standard errors of the OLS regression have been corrected for heteroskedasticity. Specification 1 shows that relation between the skill premium and marital sorting is positive and significant at the 1% level. Thus, our first empirical test agrees with the basic prediction of our theory of a positive correlation between these two variables. The next panel of Table 3 uses the skill indicator measure of the skill premium. Specification 1 again gives a positive and significant relation between the skill premium and marital sorting.

Figures 3 and 4 show the data used in the regressions of Tables 3(a) and 3(b). It is clear from these figures that Latin American countries tend to have a greater degree of inequality than the rest of our sample. One possible interpretation of this finding is that the Latin American countries are in a high inequality-high sorting steady state whereas the rest of our sample (predominantly European countries) are in a low inequality-low sorting steady state with the variation within these subsamples being explained by country-specific factors (e.g., labor-market institutions, education and tax policy, credit markets, etc.).

To make sure that our results are not driven by some factor other than sorting that is common to Latin American countries, we introduce a Latin American dummy into our regressions. As can be seen in specification 2 in Table 3, sorting is still significant in both panels, at 1% significance level with wage ratio and at

5% significance level with the skill-indicator measure. We further explore this issue by examining the relationship between inequality and sorting within the two subsamples — the LIS (including Britain) and the Latin American countries. This is done in Table 4. The relation between the skill premium and marital sorting is positive and significant in each subsample separately (though the significance varies from the 1% to the 10% levels).

Although our two measures of the skill premium have clear counterparts in our model and hence are easy to interpret, both of these measures depend on our definition of being skilled. Since this definition, i.e. going beyond high school, can be considered rather arbitrary we would like to come up with a measure that does not depend on our particular cutoff. As a widely used measure of returns to schooling, we use the Mincer coefficient as an alternative measure of the skill premium to avoid this problem. The Mincer coefficient is the coefficient on years of schooling in the following regression:[25]

$$\log(e_i) = b_0 + b_1 s_i + b_2(age - s_i - 6) + b_3(age - s_i - 6)^2 + \varepsilon_i.$$

We estimate these regression for all husbands aged 30-60 in our samples, as we did with our skill indicator measure. The fourth column of Table 1 reports our estimates of Mincer coefficient for each country. As shown in Table 2, this measure is highly correlated (over 0.8) with each of our previous two measures of the skill premium.

Table 5 show the results of using our Mincer coefficients as a measure of inequality. The data used in this regression is shown in Figure 5. In both specifications, i.e. both with and without a dummy variable for Latin American countries, the estimated relation between marital sorting and the skill premium is significant (at the 1% level without the Latin American dummy and at 5% with it). Hence, the positive correlation we found between marital sorting and the skill premium is not driven by the particular education cut-off level we chose to represent skilled versus unskilled. This gives us much greater confidence in our results as both the sample correlation measure of sorting and the Mincer coefficient measure of the skill premium depend only on our mapping of reported education to years of education. This mapping is less controversial than deciding what it takes to be

[25]These measures will differ from standard Mincer coefficients because we do not control for self-selection bias, and because we estimate the equation on husbands, rather than all working-age males. Nevertheless, our measures are quite strongly correlated (0.60) with the measures tabulated in Bils and Klenow (2000).

skilled in each country. We next examine whether the way education variables are reported and how we assign years of education affect our results.

As we have noted previously, some LIS countries report years of education whereas some report only the highest formal level attained, such as high-school diploma or undergraduate degree. As a result, for some countries the years of education or skilled category includes only those who have completed college or the appropriate degree and excludes those who have not obtained the pertinent degree but may have progressed beyond high school. In order to check whether this feature of our data affects our results, the regressions in Table 5 include a dummy variable that takes the value of 1 for countries which report the finer classifications of the data and zero otherwise. As specifications 3 and 4 in Table 5 show inclusion of this variable does not change the effect of sorting on inequality and that once we add the Latin American dummy, the new dummy variable has no additional explanatory power and the adjusted R-squared is slightly lower. This is not surprising since all Latin American countries report years of schooling which makes this new dummy variable and Latin American dummy very highly correlated.

A different concern is that although we have examined each country's education system to understand how it progresses, the actual number of years of schooling that we assign to each attainment level may affect our measure of marital sorting. A possible check is to use the rank correlation between years of schooling of husbands and wives as an alternative measure of sorting. As shown in Table 2, the rank correlation measure of sorting and sample correlation measure are highly correlated (0.96). Table 6 shows the results. As with our previous measure of sorting, we get a positive and significant relation between sorting and inequality. Sorting is again significant at the 1% level without the Latin American dummy and at 5% (wage ratio and skill indicator) and 10% level (Mincer) with the latter.

Finally, note that our analysis so far has been based on inequality in annual incomes. A better measure, were it available, would be that in expected lifetime incomes, as presumably that is what an individual is thinking about in making a tradeoff between quality and income across matches. In the absence of panel data, we cannot observe lifetime labor incomes. We can, however, create crude measures based on projections of lifetime income using the observations on older cohorts to predict the future income of the young.[26] Our simplest measure does

[26] This measure of lifetime income differs from the true measure in so far as the age-income profile varies over time.

this by dividing the life-cycle into 5-year intervals, from 25-30 up to 60-65, then computing average labor income over 5-year intervals for skilled and unskilled individuals separately. We take the present value of the predicted income profiles as the measure of lifetime labor income assuming an annual discount factor of 0.95.[27] The ratio of these lifetime income measures constitutes our fourth measure of the skill premium. The results of using this lifetime measure instead of the annual measures of inequality are given in Table 7. Sorting is again significant at 1% significance both with and without the Latin American dummy.

As a further robustness check, we also compute an analogous measure of lifetime income that controls for age variation within cohorts. We estimate the following equation:

$$y_{it} = \beta_0 + \beta_1 a_{it} + \beta_2 a_{it}^2 + \beta_3 a_{it}^3 + \gamma_0 S_i + \gamma_1 S_i a_{it} + \gamma_2 S_i a_{it}^2 + \gamma_3 S_i a_{it}^3,$$

where S_i is the indicator for being skilled and a is age. We then compute predicted income for each year for each educational class, and as before, take the ratio of the present value of the predicted income profiles as the measure of lifetime labor income inequality. This measure is highly correlated with the first measure, and the results are essentially the same as in Table 7.

3.4. Robustness

Our model abstracts from differences between men and women, both with respect to educational achievement and wage inequality. In our empirical analysis so far, we have ignored any differences that might exist between the skill premium for men and for women, and used the skill premium for husbands as our measure of inequality. We now repeat our regressions for the same sample as before but using the ratio of skilled to unskilled wives' labor earnings as our measure of inequality. Because a large proportion of women in some countries have missing values for income, the selection bias effect is much stronger than for men. Nevertheless, our measures of male and female income ratios turns out to be highly correlated (0.95). Table 8 shows that the relation between marital sorting and female wage inequality using the wage ratio measure is positive and significant at the 1% level over the entire sample, though no longer robust to inclusion of a dummy for Latin America. This suggests that female labor supply decisions might play an important role in determining the degree of inequality among females.

[27] We exclude higher ages because some of the age-country-skill cells are empty for particular countries.

We also examine how the level of financial development might affect the relation between marital sorting and the skill premium. Our model predicts that if two countries have the same level of marital sorting but differ in how binding borrowing constraints are, the country with better credit markets should exhibit less inequality. In Table 9, we introduce financial depth, measured by the M2/GDP ratio, as an additional regressor.[28] The dependent variable is the wage ratio measure of the skill premium. The first two panels of Table 9 reproduces the results from Table 3. Specification 3 shows that financial depth has negative and significant effect on the skill premium. The effect of sorting on the skill premium is positive and still significant. When we introduce a dummy variable for Latin American countries, however, the financial depth variable becomes insignificant (although it still has the right sign). This simply reflects the fact that our measure of financial depth differs systematically between Latin American countries and the rest of our sample.

Finally, as both the skill premium and marital sorting are endogenously determined variables in our model, one could in principle use an instrumental variables approach for either variable (recall that the regression in principle could be run in either direction). Our efforts in this direction were on the whole unsuccessful as described below.

As a possible instrument for the skill premium, we examined capital per worker. Although this variable is strongly correlated with the skill premium in our full sample, it does not capture the variation in the skill premium beyond its variation between Latin American countries and the rest of our sample. In order to avoid this problem we also tried to use capital per worker as an instrument for the skill premium within our two subsamples (Latin America and LIS). It turned out to be not a good instrument, since it is only weakly correlated with skill premium within these subsamples.

Finding an appropriate instrument for marital sorting is even harder. One possible candidate is the ratio of women to men in the marriage market. A possible measure of this is the ratio of women within a certain age bracket (say 25-50) to men within a slightly different bracket (say 36-45) which may plausibly capture the group of women from which men find spouses. This variable has an average value of 2.28 and has a standard deviation of 0.36; the variation arises from differences in the age distribution of the population.

[28] M2/GDP data is for 1994. The year 1994 was chosen to be able to have data for all countries in a year around the survey dates (as the data on ex-communist countries is only available after 1993). The results with using longer averages for other countries are very similar.

We find that this ratio is in fact positively correlated with sorting (Pearson coefficient = 0.53), and that the value of sorting predicted by the sex ratio does an excellent job of predicting the wage ratio measure of the skill premium across countries (R-squared = 0.47). However, as an instrument for sorting, the sex ratio suffers from exactly the same problems as capital per worker does as an instrument for the skill premium; due to systematic differences between Latin America and the rest of the sample, inclusion of a dummy variable for Latin America renders the effect of the instrument on inequality insignificant.

3.5. Per Capita Income, Skilled Population and Sorting

We now turn to an examination of another prediction of our model: the existence of a negative relation between marital sorting and per capita income across countries. Note that our model implies that an economy in a steady state with a high degree of marital sorting will have a large skill premium and in particular a large fraction of individuals facing credit constraints in their education decisions. Consequently, ceteris paribus, we expect economies with similar technologies but with greater sorting to have lower per capita income as their level of human capital will be further below the efficient level. The observed relation between marital sorting and per capita income is shown in Figure 6. The per capita income measure is real GDP per capita in 1997 from World Bank Global Development Network Growth Database.[29] Table 10 shows the regression results for a specification in which the dependent variable is per capita income and the explanatory variable is the sample correlation measure of marital sorting with and without a Latin American dummy. The relation is significant and negative for both specifications.

We can study the hypothesis developed above more closely by examining whether there exists a negative relationship (as predicted by our model when countries have identical technologies) between the fraction of the population that is skilled and the degree of sorting. This is done in Table 11, where we report the relationship between the proportion of the (male and female) population that is skilled in our sample and marital sorting.[30] In both cases, the relations are negative and significant as our model predicts.

[29] The data for Germany is from 1992.

[30] We run this regression both for skilled men as a fraction of the male population and skilled women as a fraction of the female population. These two measures are in any case highly correlated (0.94).

4. Conclusion

This paper has examined the relationship between marital or household sorting and income inequality. Using a simple model in which individuals make decisions over whether to become skilled or unskilled, about with whom to match, how much to consume, and the number of children to have, we find that there is a positive relationship between sorting and inequality (between skilled and unskilled workers). In particular, whether at a point in time, or across steady states, economies with greater skill premia should also display a greater degree of sorting. Our model also predicts that economies with greater skill premia should have greater fertility differentials, and (given identical technologies) economies with greater sorting should have lower per capita income and smaller fractions of skilled workers.

Our empirical work, based on household surveys for 34 countries, and using various measures of inequality and marital sorting, supports our central prediction of a positive relationship between sorting and inequality across countries. We also find evidence in favor of a negative relationship between sorting and per capita income as well as between sorting and the fraction of the population that is skilled.

It should be noted that our story of greater pickiness with respect to household partners in the face of an increased skill premium is of course not the only one compatible with a positive correlation between these two variables. An alternative story, with similar mechanics, would be of individuals sorting more into communities or schools in response to greater inequality (say, in response to fear of more crime). This could then lead to fewer opportunities to interact between individuals of different skill groups and consequently to a greater correlation of spouses in education. We do not see this mechanism as being very different. Once again, private decisions (e.g., where to live, where to go to school, who to marry) would have important social consequences as a result of borrowing constraints.

There are many directions in which this work could be extended. We have abstracted from several issues, each of which are of interest in their own right. First, we have ignored differences between men and women. An alternative formulation of our model would be to have parents care about the quality and quantity of their children and for parental time and education to be a factor in producing quality (perhaps by lowering the cost of the children becoming skilled). Thus, a parent who stayed at home and took care of the children would contribute to household utility by increasing the quality of their offspring. If, because of childbearing costs this were predominantly the woman, men would still wish to match with more educated women either because of their earning potential (as in the model) or because of the increased quality of the children. Thus, a major topic we wish to investigate (theoretically and empirically) is the relationship among sorting, female wage inequality and male wage inequality.[31] This would also tie in with another set of issues that we have chosen to ignore–that of household bargaining, the option to remain single and the possibility of divorce. Another avenue to explore is the importance of bequests relative to education in the intergenerational transmission of inequality. Lastly, it would be interesting to examine the role of public policy (education subsidies and welfare policy) in interacting with sorting and inequality. We plan to study several of these issues in future work.

[31]See Galor and Weil (1996) for a model in which exogenous differences between women and men leads to a large gap between the wages of these at low levels of capital, which is then reduced as capital accumulates. They use this model to help explain the demographic transition.

References

[1] Aiyagari, S. Rao, J. Greenwood, and N. Guner. "On the State of the Union." Journal of Political Economy, 108 (2000), 213-244.

[2] Becker, Gary S. and N. Tomes, "An Equilibrium Theory of the Distribution of Income and Intergenerational Mobility," Journal of Political Economy, 87 (1979), 1153-1189.

[3] _ _ _ _ _ _ _ _, "Human Capital and the Rise and Fall of Families," Journal of Labor Economics, 4 (1986), S1-S39.

[4] Bénabou, Roland, "Heterogeneity, Stratification and Growth," American Economic Review, 86 (1996), 584-609.

[5] Bergstrom, Theodore C., "A Survey of Theories of The Family," in Handbook of Population and Family Economics, Edited by M. R. Rosenzweig and O. Stark, Elsevier Science, 1997.

[6] Bils, Mark and P. J. Klenow, "Does Schooling Cause Growth?" American Economic Review, 90 (2000), 1160-1183.

[7] Boulier, Bryan L. and M. R. Rosenzweig, "Schooling, Search, and Spouse Selection: Testing Economic Theories of Marriage and Household Behavior," Journal of Political Economy, 92 (1984), 712-732.

[8] Burdett, Kenneth and M. G. Coles, "Marriage and Class," Quarterly Journal of Economics, 112, (1997), 141-168.

[9] _ _ _ _ _ _ _ _ _, "Transplants and Implants: The Economics of Self-Improvement," mimeo, (1999).

[10] Cole, Harold L., G. Mailath, and A. Postlewaite, "Social Norms, Savings Behavior and Growth," Journal of Political Economy, 100 (1992), 1092-1125.

[11] Dahan, Momi and D. Tsiddon, "Demographic Transition, Income Distribution and Economic Growth," Journal of Economic Growth, 3 (1998), 29-52.

[12] Dahan, Momi and A. Gaviria, "Sibling Correlations and Intergenerational Mobility in Latin America," mimeo, (1999).

[13] Durlauf, S., “A Theory of Persistent Income Inequality,” Journal of Economic Growth, 1 (1995), 75-93.

[14] Fernández, Raquel, “Education, Segregation, and Marital Sorting: Theory and an Application to UK Data,” working paper, (2001).

[15] Fernández, Raquel and C. A. Pissarides, “Searching for a Partner,” mimeo, (2001).

[16] Fernández, Raquel and R. Rogerson, “Public Education and Income Distribution: A Dynamic Quantitative Evaluation of Education Finance Reform,” American Economic Review, 88 (1998), 813-833.

[17] _ _ _ _ _ _ _ _ _ , “Sorting and Long-Run Inequality,” Quarterly Journal of Economics, forthcoming.

[18] Galor, Oded and J. Zeira, “Income Distribution and Macroeconomics,” Review of Economic Studies, 60 (1993), 35-52.

[19] Galor, Oded and D. Weil, “The Gender Gap, Fertility and Growth,” American Economic Review, 86 (1996), 374-387.

[20] Greenwood, Jeremy, N. Guner and J. Knowles, “More on Marriage, Fertility and the Distribution of Income,” mimeo, (1999).

[21] Knowles, John, “Measuring Lifetime Income Inequality,” mimeo, (1999).

[22] Kremer, Michael, “How Much does Sorting Increase Inequality,” Quarterly Journal of Economics, 112 (1997), 115-139.

[23] Kremer, Michael and D. Chen, “Income Distribution Dynamics with Endogenous Fertility,” American Economic Review Papers and Proceedings, 89 (1999), 155-160.

[24] Lam, David, “Marriage Markets and Assortative Mating with Household Public Goods: Theoretical Results and Empirical Implications,” Journal of Human Resources, 23 (1988), 462-487.

[25] Ljungqvist, Lars, “Economic Development: The Case of a Missing Market for Human Capital,” Journal of Development Economics, 40 (1993), 219-239.

[26] Loury, Glenn C. "Intergenerational Transfers and the Distribution of Earnings." Econometrica, 49 (1981), 843-867.

[27] Mare, Robert D. "Five Decades of Educational Assortative Mating," American Journal of Sociology, 56, (1991), 15-32.

[28] Owen, Ann and D. Weil, "Intergenerational Earnings Mobility, Inequality and Growth," Journal of Monetary Economics, 41 (1998), 71-104.

[29] Perotti, Roberto, "Growth, Income Distribution, and Democracy: What the Data Say," Journal of Economic Growth, 1 (1996), 149-187.

[30] Pissarides, Chris, "From School to University: The Demand for Post-Compulsory Education in Britain," Economic Journal, 92 (1982), 654-667.

[31] Regalia, Ferdinando and J. V. Rios-Rull. "What Accounts for the Increase in Single Households and the Stability in Fertility?" mimeo (1999).

[32] Smits, Jeroen, W. Ultee, and J. Lammers, "Educational Homogamy in 65 Countries: An Explanation of Differences in Openness Using Country-Level Explanatory Variables," American Sociological Review, 63 (1998), 264-285.

[33] Weiss, Yoram. "The Formation and Dissolution of Families: Why Marry? Who Marries Whom? and What Happens upon Divorce?" in Handbook of Population and Family Economics, Edited by M. R. Rosenzweig and O. Stark, Elsevier Science, 1997.

5. Appendix A

We will now show that all terms in the first curly bracket of equation (2.17) Before doing this, as we have already argued, $\frac{\partial q^*}{\partial k}$ is strictly negative, since

$$\frac{\partial q^*}{\partial k} = \underbrace{\frac{f''}{f'(f'(1-k)+f)}}_{-} \left[\underbrace{(\widetilde{w}_s - w_u)}_{+} + f'(1+k)\underbrace{(w_u + \widetilde{w}_s - \beta)}_{+} \right] < 0, \tag{5.1}$$

where $f'(f'(1-k)+f) = \widetilde{w}_s(\widetilde{w}_s + w_u) > 0$. Hence, all we want to determine is the sign of the following expression

$$\begin{aligned} A &= \frac{f''}{(\widetilde{w}_s + w_u)w_u\widetilde{w}_s}\frac{1}{1+k}\{[(1+2kQ^* + k^2)(\widetilde{w}_s + w_u - \beta) + (2k + Q^*k^2 + Q^*)(\widetilde{w}_s + w_u)] \\ &\quad \widetilde{w}_s w_u - \beta(k+Q^*)w_u^2 - \beta(k + k^2 Q^*)\widetilde{w}_s^2\}, \end{aligned} \tag{5.2}$$

where $Q^* = Q(q^*)$.

Note that if $w_u \geq \beta$, then we are all set, since then

$$(2k + Q^*k^2 + Q^*)\widetilde{w}_s^2 w_u - \beta(k + k^2Q^*)\widetilde{w}_s > 0,$$

and

$$(2k + Q^*k^2 + Q^*)w_u^2 w_s - \beta(k + Q^*)w_u^2 > 0.$$

Therefore, we only need to take care of the case where $w_u < \beta$.

In order to show that the following expression

$$\begin{aligned} A &= \{[(1+2kQ^* + k^2)(\widetilde{w}_s + w_u - \beta) + (2k + Q^*k^2 + Q^*)(\widetilde{w}_s + w_u)]\widetilde{w}_s w_u \\ &\quad -\beta(k+Q^*)w_u^2 - \beta(k + k^2Q^*)\widetilde{w}_s^2\}, \end{aligned}$$

is positive for $w_u < \beta$, we will simply show that it is increasing in w_u and $\widetilde{w}_s$ and when evaluated at $w_u = \widetilde{w}_s = \frac{\beta}{2}$, it is non-negative (recall that $w_u \geq \frac{\beta}{2}$ and $w_u \geq \frac{\beta}{2}$). We start by showing that A is increasing in w_u for $w_u < \beta$. In order to do this, let take the derivative of A with respect to w_u to get

$$\begin{aligned} \frac{\partial A}{\partial w_u} &= (1+2kQ^* + k^2)(\widetilde{w}_s^2 + 2w_s w_u - \beta\widetilde{w}_s) \\ &\quad +(2k + Qk^2 + Q)(2\widetilde{w}_s w_u + \widetilde{w}_s^2) - 2\beta(k+Q)w_u. \end{aligned}$$

Note that this expression is increasing in $\widetilde{w}_s$ (since $\widetilde{w}_s \geq \frac{\beta}{2}$), and we can evaluate it at the limit where $\widetilde{w}_s = \frac{\beta}{2}$ to get

$$\begin{aligned}\left.\frac{\partial A}{\partial w_u}\right|_{\widetilde{w}_s=\frac{\beta}{2}} &= \beta w_u(1+2kQ+k^2+2k+Qk^2+Q-2k-2Q) \\ &\quad +\frac{\beta^2}{4}(2k+Qk^2+Q-1-2kQ-k^2).\end{aligned}$$

This expression is also increasing in w_u (since $1 \geq Q$), and hence we can also evaluate it at $w_u = \frac{\beta}{2}$ (recall that $w_u \geq \frac{\beta}{2}$) to get

$$\left.\frac{\partial A}{\partial w_u}\right|_{\widetilde{w}_s=\frac{\beta}{2}, w_u=\frac{\beta}{2}} = \beta^2\left(\frac{1}{4}+\frac{1}{2}kQ+\frac{1}{4}k^2-\frac{1}{4}Q+\frac{3}{4}Qk^2+\frac{1}{2}k\right) > 0.$$

Therefore, A is indeed increasing in w_u.

We next will show that A is increasing in $\widetilde{w}_s$. Taking the derivative with respect to $\widetilde{w}_s$ we get

$$\begin{aligned}\frac{\partial A}{\partial \widetilde{w}_s} &= (1+2kQ+k^2)(2\widetilde{w}_s w_u+\widetilde{w}_u^2-\beta w_u) \\ &\quad +(2k+Qk^2+Q)(2\widetilde{w}_s w_u+w_u^2)-2\beta(k+k^2Q^*)\widetilde{w}_s.\end{aligned}$$

Since $w_u \geq \frac{\beta}{2}$, this expression is increasing in w_u, and we can evaluate at the limit where $w_u = \frac{\beta}{2}$,

$$\begin{aligned}\left.\frac{\partial A}{\partial \widetilde{w}_s}\right|_{w_u=\frac{\beta}{2}} &= (1+2kQ+k^2)(\beta\widetilde{w}_s-\frac{\beta^2}{4}) \\ &\quad +(2k+Qk^2+Q)(\beta\widetilde{w}_s+\frac{\beta^2}{4})-2\beta(k+k^2Q)\widetilde{w}_s.\end{aligned}$$

Again, this expression is also increasing in $\widetilde{w}_s$ (note that $Q \leq 1$), therefore we can evaluate it at $\widetilde{w}_s = \frac{\beta}{2}$,

$$\left.\frac{\partial A}{\partial \widetilde{w}_s}\right|_{\widetilde{w}_s=\frac{\beta}{2}, w_u=\frac{\beta}{2}} = (1+2kQ+k^2)\frac{\beta^2}{4}+\frac{3\beta^2}{4}(2k+Qk^2+Q)-(k+k^2Q)\beta^2 > 0.$$

Thus A is increasing in $\widetilde{w}_s$ and w_u. To show that A is positive, we simply evaluate it at $\widetilde{w}_s = w_u = \frac{\beta}{2}$:

$$A|_{\widetilde{w}_s=\frac{\beta}{2},w_u=\frac{\beta}{2}} = 0 + (2k + Q^*k^2 + Q^*)\frac{\beta^3}{4} - (k + Q^*)\frac{\beta^3}{4} - (k + k^2Q^*)\frac{\beta^3}{4} = 0$$

Hence, in Equation (2.17) all terms are negative except those with $\frac{\partial\lambda}{\partial k}$.

6. Appendix B

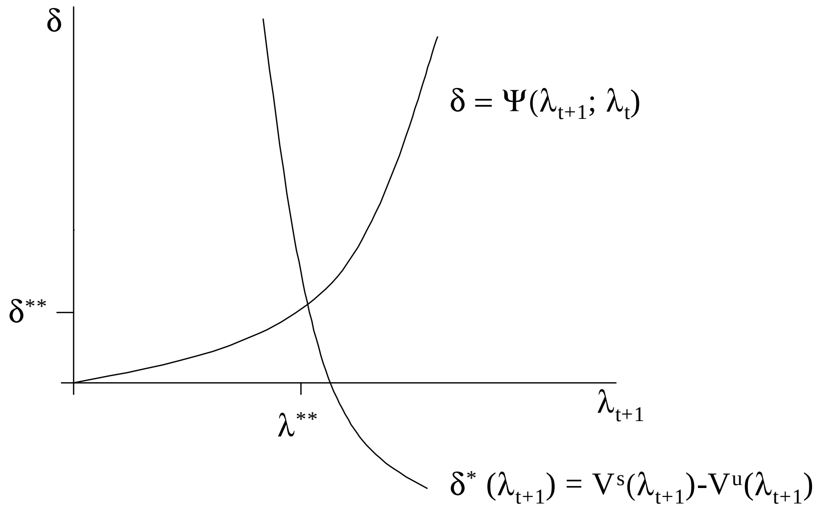
The years of households surveys used in our empirical study are given in Table B1. All surveys are nationally representative samples, except for Argentina and Uruguay for which we have only urban samples (70% of the population for Argentina and 90% for Uruguay). Table B1 also gives details of the income measures available in each survey. The income in the Latin American countries is gross monthly labor income from all sources. This definition varies across countries, but generally includes wages, income from self-employment, proprietor's income, from both primary and secondary labor activities. Some LIS countries report gross annual earnings and income and some report these net of taxes.[32] We use gross labor earnings for LIS countries whenever it is available.. The gross earnings measure for LIS countries include all forms of cash wage and salary income, including employer bonuses, 13th month bonus, etc., (gross of employee social insurance contributions/taxes but net of employer social insurance contributions/taxes). While most countries report gross earnings, the following countries report only the net earnings: France, Hungary, Italy, Poland, Russia, and Spain. Since taxation tends to be progressive in the countries we are comparing, inequality of income is likely to be higher than reported in those countries for which pre-tax income is not reported. We do not adjust income measures in LIS or IDB for hours worked or weeks worked in order to arrive at a measure of total income, including leisure. This is because few countries collect hours or weeks series, and some of those that do collect them, such as Slovakia or Spain, use discrete codes rather then report actual levels.

Education in the Latin American data is reported as total years of schooling. For the LIS countries the education units are quite idiosyncratic. We attempt to

[32] Some LIS countries are excluded because they do not report all of the variables required for the analysis. Ireland and Austria do not report individual labor income. Ireland also does not report the education of the spouse in the household sample. Education variables are not available for Switzerland.

standardize the LIS education variable by converting the reported units to years of education. We define as skilled all agents who went beyond high school education. For some of the Latin-American data, this coincides with indicator variables for higher education, as a few of these countries report attainment in addition to years of education. For the other Latin-America countries, this indicator is constructed using the standard age-grade progression for that country. Thus, skilled workers in Costa Rica, for example, are those with more than 11 years of education, while in Mexico, they are those with more than 12 years. For Britain we would have preferred to define as skilled any individual with at least 2A levels passes (as in Fernández (2001) or Pissarides (1982)), but as the data did not permit us to distinguish among individuals with different number of A levels, we instead categorized them all as unskilled. Our results are robust to categorizing them all as skilled instead. Table B2 reports our mapping of education measures into years of schooling and into an indicator for high school completions. For most countries, we were able to compare the percentage of adults with education beyond the high-school level to published sources, and to reconcile our statistics with the previously published numbers.

Figure 1


δ
δ = Ψ(λt+1; λt)
δ**
λ**
λt+1
δ* (λt+1) = Vs(λt+1)-Vu(λt+1)

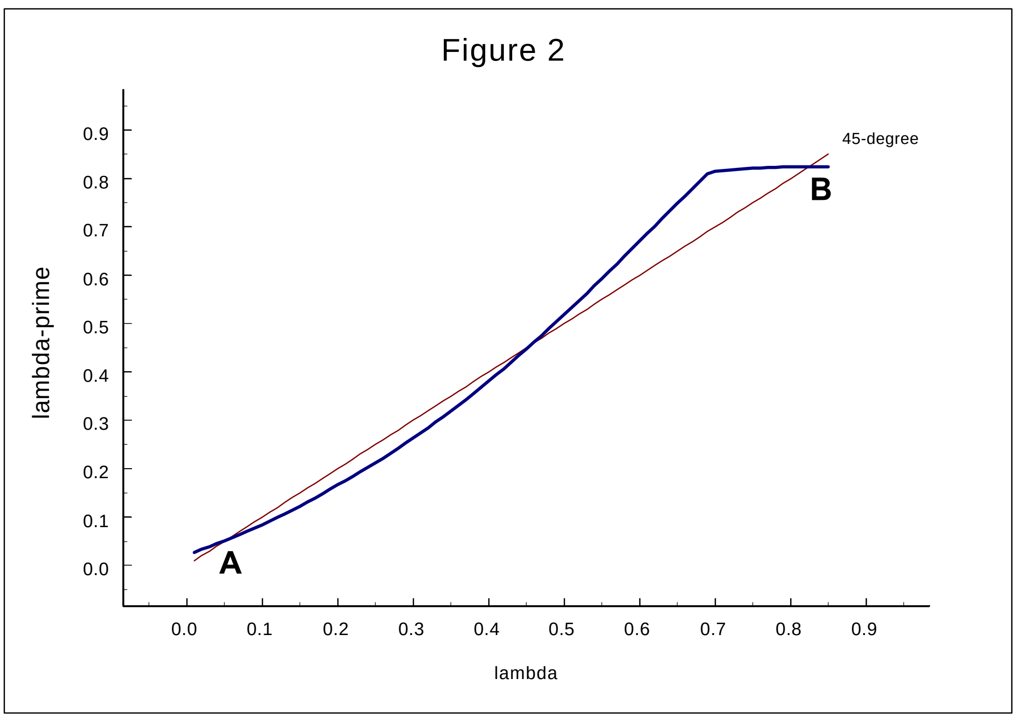


Figure 2
45-degree
B
A
lambda-prime
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
0.0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9
lambda

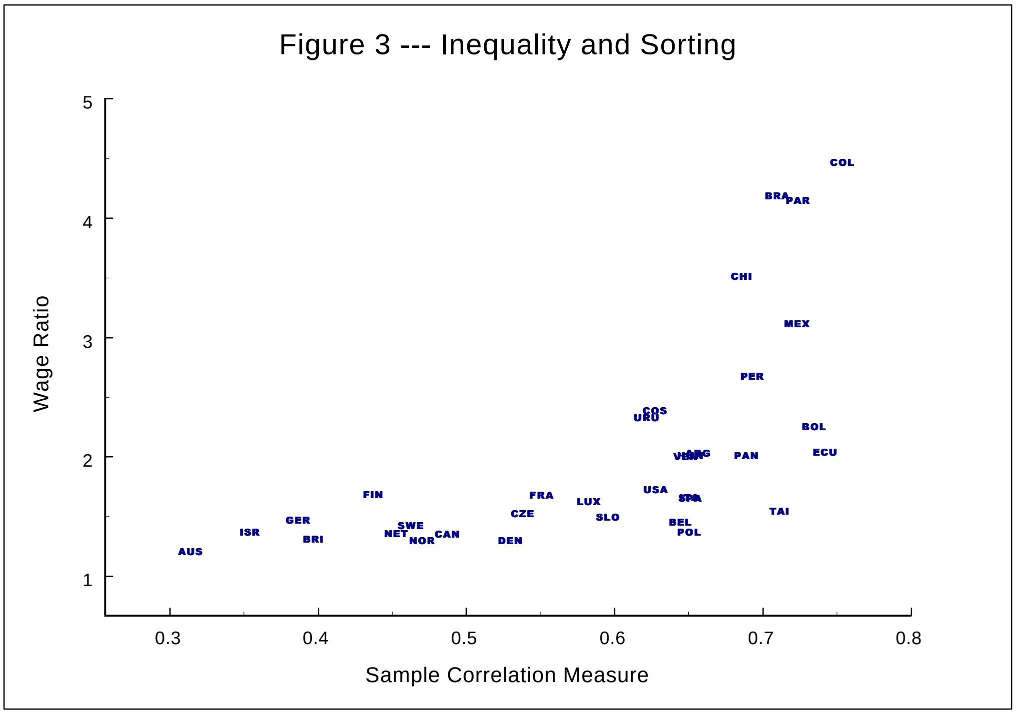


Figure 3 --- Inequality and Sorting
Wage Ratio
Sample Correlation Measure
5
4
3
2
1
0.3
0.4
0.5
0.6
0.7
0.8
COL
BRA
PAR
CHI
MEX
PER
COS
URU
BOL
ARG
VEN
PAN
ECU
FIN
FRA
LUX
USA
TAI
GER
CZE
SLO
BEL
ISR
SWE
NET
CAN
POL
BRI
NOR
DEN
AUS

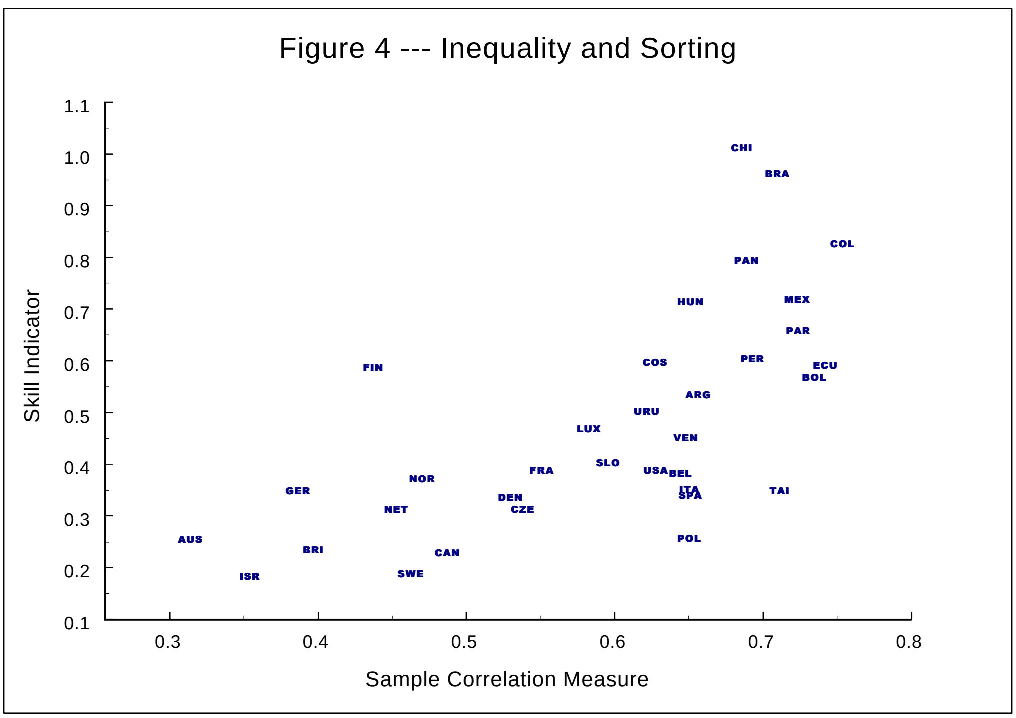


Figure 4 --- Inequality and Sorting
Skill Indicator
Sample Correlation Measure
1.1
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.3
0.4
0.5
0.6
0.7
0.8
CHI
BRA
COL
PAN
HUN
MEX
PAR
FIN
COS
PER
ECU
BOL
ARG
URU
LUX
VEN
FRA
SLO
USA
BEL
NOR
GER
ITA
SPA
TAI
DEN
NET
CZE
AUS
POL
BRI
CAN
ISR
SWE

Figure 5 --- Inequality and Sorting

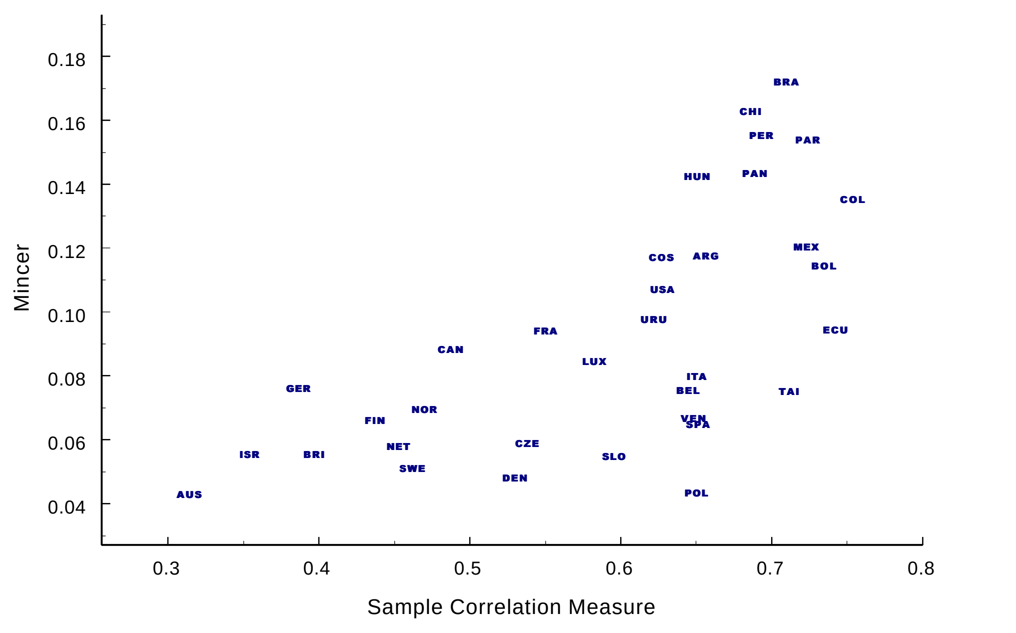


Figure 6 --- GDP per capita and Sorting

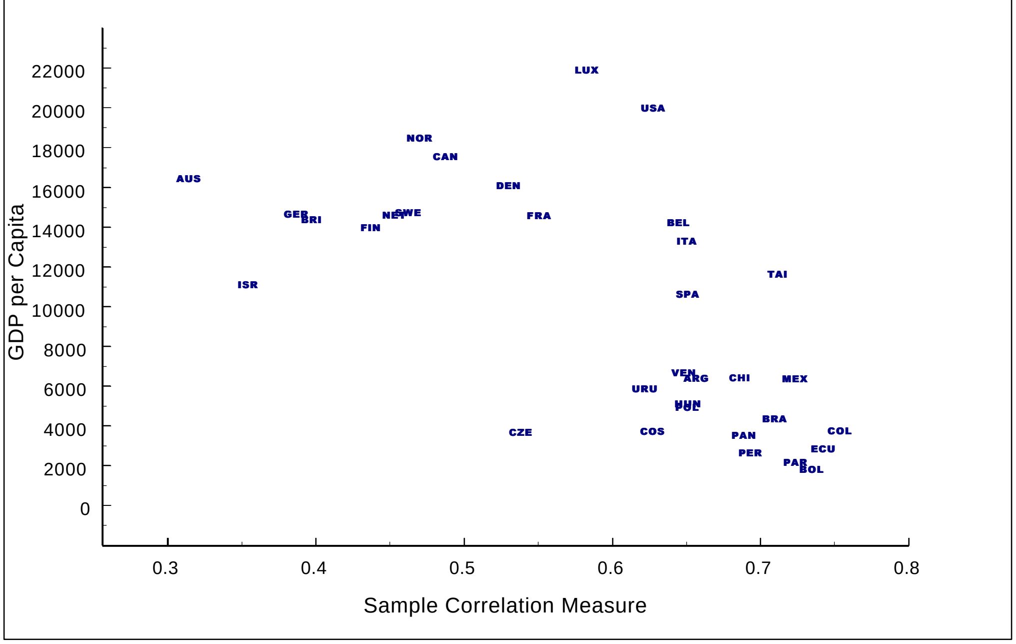


Table 1: Estimates of skill premium and marital sorting

Country	Statistic	Skilled Share	Skill Premium			Marital Sorting	
		Husbands	Wage Ratio	Skill Indicator	Mincer	Sample Correlations	Rank Correlations
Argentina	mean	0.259	2.048	0.538	0.118	0.657	0.654
	std.	(0.438)	(0.096)	(0.022)	(0.003)	.	
Australia	mean	0.298	1.224	0.258	0.043	0.314	0.282
	std.	(0.457)	(0.016)	(0.039)	(0.005)	.	
Belgium	mean	0.318	1.469	0.385	0.076	0.645	0.622
	std.	(0.466)	(0.105)	(0.031)	(0.005)	.	
Bolivia	mean	0.184	2.268	0.570	0.115	0.734	0.711
	std.	(0.388)	(2.154)	(0.065)	(0.007)	.	
Brasil	mean	0.109	4.202	0.964	0.172	0.710	0.687
	std.	(0.312)	(2.066)	(0.021)	(0.002)	.	
Britain	mean	0.488	1.324	0.238	0.056	0.397	0.386
	std.	(0.500)	(0.161)	(0.027)	(0.005)	.	
Canada	mean	0.585	1.368	0.232	0.089	0.487	0.452
	std.	0.493	(0.288)	(0.017)	(0.002)	.	
Chile	mean	0.148	3.525	1.015	0.163	0.686	0.693
	std.	0.355	(0.273)	(0.022)	(0.002)	.	
Colombia	mean	0.148	4.478	0.829	0.136	0.753	0.745
	std.	(0.356)	(20.500)	(0.028)	(0.002)	.	
Costa Rica	mean	0.157	2.401	0.601	0.118	0.627	0.586
	std.	(0.364)	(0.734)	(0.047)	(0.005)	.	
Czech	mean	0.130	1.539	0.316	0.059	0.538	0.523
	std.	(0.336)	(0.221)	(0.019)	(0.002)	.	
Denmark	mean	0.260	1.310	0.339	0.049	0.530	0.447
	std.	(0.439)	(0.001)	(0.041)	(0.005)	.	
Ecuador	mean	0.275	2.054	0.594	0.095	0.742	0.748
	std.	(0.447)	(0.013)	(0.039)	(0.004)	.	
Finland	mean	0.187	1.701	0.591	0.067	0.437	0.449
	std.	(0.390)	(0.033)	(0.060)	(0.008)	.	
France	mean	0.218	1.692	0.391	0.095	0.550	0.538
	std.	(0.413)	(0.033)	(0.037)	(0.005)	.	
Germany	mean	0.246	1.487	0.352	0.077	0.387	0.312
	std.	(0.431)	(0.085)	(0.032)	(0.005)	.	
Hungary	mean	0.175	2.027	0.717	0.143	0.651	0.622
	std.	(0.381)	(0.019)	(0.072)	(0.012)	.	
Israel	mean	0.426	1.383	0.186	0.056	0.354	0.482
	std.	(0.495)	(0.065)	(0.032)	(0.004)	.	
Italy	mean	0.104	1.670	0.354	0.080	0.650	0.636
	std.	(0.305)	(0.439)	(0.033)	(0.003)	.	
Luxemb	mean	0.098	1.641	0.472	0.085	0.583	0.557
	std.	(0.300)	(0.142)	(0.042)	(0.005)	.	
Mexico	mean	0.174	3.131	0.721	0.121	0.723	0.704
	std.	(0.379)	(0.711)	(0.039)	(0.004)	.	
Nether	mean	0.253	1.370	0.316	0.058	0.453	0.438
	std.	(0.435)	(0.138)	(0.031)	(0.004)	.	
Norway	mean	0.269	1.312	0.376	0.070	0.470	0.430
	std.	(0.443)	(0.035)	(0.044)	(0.008)	.	
Panama	mean	0.193	2.022	0.797	0.144	0.689	0.702
	std.	(0.395)	(2.927)	(0.063)	(0.007)	.	
Paraguay	mean	0.090	4.163	0.661	0.154	0.724	0.665
	std.	(0.286)	(2.919)	(0.104)	(0.009)	.	
Peru	mean	0.234	2.688	0.607	0.156	0.693	0.698
	std.	(0.423)	(0.683)	(0.082)	(0.010)	.	
Poland	mean	0.105	1.381	0.260	0.044	0.650	0.634
	std.	(0.306)	(0.025)	(0.034)	(0.004)	.	
Slovakia	mean	0.152	1.511	0.407	0.055	0.595	0.615
	std.	(0.360)	(0.023)	(0.016)	(0.002)	.	
Spain	mean	0.210	1.668	0.343	0.065	0.651	0.713
	std.	(0.408)	(0.107)	(0.043)	(0.002)	.	
Sweden	mean	0.326	1.439	0.191	0.051	0.462	0.466
	std.	(0.469)	(0.149)	(0.047)	(0.004)	.	
Taiwan	mean	0.258	1.559	0.352	0.075	0.711	0.717
	std.	(0.438)	(0.026)	(0.016)	(0.002)	.	
Uruguay	mean	0.235	2.339	0.505	0.098	0.622	0.597
	std.	(0.424)	(0.260)	(0.035)	(0.004)	.	
USA	mean	0.588	1.743	0.391	0.108	0.627	0.606
	std.	(0.492)	(0.222)	(0.014)	(0.002)	.	
Venezuela	mean	0.156	2.015	0.454	0.067	0.648	0.642
	std.	(0.363)	(0.445)	(0.047)	(0.005)	.	.
Sample	mean	0.237	2.034	0.480	0.093	0.593	0.581
	std.	(0.126)	(0.883)	(0.215)	(0.038)	(0.122)	(0.126)

Table 2: Correlations

	Skill Premium				Sorting	
	Wage Ratio	Skill Indicator	Mincer	Lifetime Income ratio	Sample Correlation	Rank Correlation
Skill Premium						
Wage Ratio	1.000					
Skill Indicator	0.835	1.000				
	(0.000)					
Mincer	0.803	0.879	1.000			
	(0.000)	(0.000)				
Lifetime Income ratio	0.923	0.851	0.830	1.000		
	(0.000)	(0.000)	(0.000)			
Sorting						
Sample Correlation	0.632	0.655	0.658	0.593	1.000	
	(0.000)	(0.000)	(0.000)	(0.000)		
Rank Correlation	0.586	0.610	0.605	0.545	0.958	1.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	

Significance levels are shown in paranthesis

Table 3: Regression of Skill Premium on Marital Sorting (Sample Correlation Measure)

	Dependent Variable			
	(a) Wage Ratio		(b) Skill Indicator	
Explanatory Variable	1	2	1	2
Constant	-0.664	0.603	-0.202	0.084
	(0.529)	(0.312)*	(0.116)*	(0.116)
Sorting	4.552	1.720	1.150	0.512
	(1.029)***	(0.595)***	(0.211)***	(0.222)**
LA Dummy		1.078		0.243
		(0.234)***		(0.068)***
Adjusted R-square	0.3803	0.582	0.429	0.581

Standard errors reflect Eicker-White correction for heteroscedasticity.
* Significant at 10%.
** Significant at 5%.
*** Significant at 1%

Table 4: Regression of Skill Premium on Marital Sorting (Sample Correlation Measure)

	Dependent Variable			
	(a) Wage Ratio		(b) Skill Indicator	
Explanatory Variable	Latin America	Rest	Latin America	Rest
Constant	-4.366	1.006	-0.408	0.133
	(3.456)	(0.151)***	(0.509)	(0.119)
Sorting	10.447	0.960	1.572	0.420
	(5.220)*	(0.305)***	(0.743)*	(0.228)*
Adjusted R-square	0.175	0.291	0.078	0.1025

Standard errors reflect Eicker-White correction for heteroscedasticity.
* Significant at 10%.
** Significant at 5%.
*** Significant at 1%

Table 5: Regression of Mincer Coefficient on Marital Sorting (Sample Correlation Measure)

	Dependent Variable Mincer Coefficient			
Explanatory Variable	1	2	3	4
Constant	-0.028	0.019	-0.024	0.020
	(0.017)	(0.019)	(0.016)	(0.019)
Sorting	0.204	0.098	0.180	0.098
	(0.033)***	(0.040)**	(0.032)***	(0.041)**
Education Dummy			0.015	-0.000
			(0.009)	(0.009)
LA Dummy		0.040		0.040
		(0.013)***		(0.014)***
Adjusted R-square	0.415	0.562	0.431	0.547

Standard errors reflect Eicker-White correction for heteroscedasticity.
* Significant at 10%.
** Significant at 5%.
*** Significant at 1%

Table 6: Regression of Skill Premium on Marital Sorting (Rank Correlation Measure)

	Dependent Variable					
	(a) Wage Ratio		(b) Skill Indicator		(c) Mincer	
Explanatory Variable	1	2	1	2	1	2
Constant	-0.349	0.833	-0.125	0.144	-0.013	0.033
	(0.491)	(0.275)***	(0.121)	(0.094)	(0.019)	(0.018)*
Sorting	4.100	1.310	1.041	0.407	0.182	0.074
	(0.995)***	(0.537)**	(0.225)***	(0.185)**	(0.037)***	(0.038)*
LA Dummy		1.148		0.261		0.044
		(0.244)***		(0.063)***		(0.012)***
Adjusted R-square	0.323	0.570	0.353	0.566	0.346	0.568

Standard errors reflect Eicker-White correction for heteroscedasticity.
* Significant at 10%.
** Significant at 5%.
*** Significant at 1%

Table 7: Regression of Skill Premium on Marital Sorting (Sample Correlation Measure)

	Dependent Variable Lifetime Income Ratio	
Explanatory Variable	1	2
Constant	-0.363 (0.479)	0.700 (0.292)**
Sorting	4.003 (0.951)***	1.602 (0.567)***
LA Dummy		0.916 (0.264)***
Adjusted R-square	0.330	0.485

Standard errors reflect Eicker-White correction for heteroscedasticity.
* Significant at 10%.
** Significant at 5%.
*** Significant at 1%

Table 8: Regression of Female Skill Premium on Marital Sorting (Sample Correlation Measure)

	Dependent Variable Female Wage Ratio	
Explanatory Variable	1	2
Constant	-0.237 (0.498)	0.950 (0.351)**
Sorting	3.789 (0.947)***	1.135 (0.671)
LA Dummy		1.010 (0.226)***
Adjusted R-square	0.321	0.538

Standard errors reflect Eicker-White correction for heteroscedasticity.
* Significant at 10%.
** Significant at 5%.
*** Significant at 1%

Table 9: Regression of Skill Premium on Marital Sorting (Sample Correlation Measure)

	Dependent Variable Wage Ratio			
Explanatory Variable	1	2	3	4
Constant	-0.664 (0.529)	0.603 (0.312)***	0.373 (0.418)	0.752 (0.281)**
Sorting	4.552 (1.029)***	1.720 (0.595)***	3.770 (0.816)***	1.942 (0.687)***
Financial Depth			-0.011 (0.003)***	-0.004 (0.003)
LA Dummy		1.078 (0.234)***		0.885 (0.261)***
Adjusted R-square	0.3803	0.582	0.501	0.579

Standard errors reflect Eicker-White correction for heteroscedasticity.
* Significant at 10%.
** Significant at 5%.
*** Significant at 1%

Table 10: Regression of GDP per Capita on Marital Sorting (Sample Correlation Measure)

	Dependent Variable GDP per Capita	
Explanatory Variable	1	2
Constant	28276.15 (2977.60)***	20454.97 (3077.80)***
Sorting	-31011.62 (4828.80)***	-13520.90 (6639.63)**
LA Dummy		-6654.80 (1929.27)***
Adjusted R-square	0.391	0.558

Standard errors reflect Eicker-White correction for heteroscedasticity.
* Significant at 10%.
** Significant at 5%.
*** Significant at 1%

Table 11: Regression of Fraction of Skilled Population on Marital Sorting (Sample Correlation Measure)

	Dependent Variable			
	(a) Skilled Male Population		(b) Skilled Female Population	
Explanatory Variable	1	2	1	2
Constant	0.514 (0.089)***	0.486 (0.111)***	0.485 (0.088)***	0.463 (0.110)***
Sorting	-0.468 (0.138)***	-0.406 (0.213)*	-0.478 (0.135)***	-0.430 (0.209)**
LA Dummy		-0.024 (0.053)		-0.018 (0.051)
Adjusted R-square	0.183	0.162	0.197	0.175

Standard errors reflect Eicker-White correction for heteroscedasticity.
* Significant at 10%.
** Significant at 5%.
*** Significant at 1%

TABLE B1: Survey Information

Country	Year	Name	Agency	Number of Households	Coverage
Argentina	1996	Encuesta permanente de hogares	Instituto nacional de Estadistica y Censos	3369	Greater Buenos Aires
Australia	1994	Australian Income and Housing Survey	Australian Bureau of Statistics	7441	National
Belgium	1992	Panel Survey of the Centre for Social Policy	Centre for Social Policy	3821	National
Bolivia	1997	Encuesta Nacional de Empleo	Instituto nacional de Estadistica y Censos	8461	National
Brasil	1996	Pequiso Nacional por Amostra de Domicilios	Fundaçao Instituto Brasileiro de Geografia e Estatistica	84947	National
Britain	1997	British Household Panel Study, Wave G	Institute for Social and Economic Research	4384	National
Canada	1994	Survey of Consumer Finances	Statistics Canada	39039	National
Chile	1996	Encuesta Nacional de Empleo	Instituto Nacional de Estadistica	30953	National
Colombia	1997	Encuesta Nacional de Hogares	Departamento Administrativo Nacional de Estadistica	31264	National
Costa Rica	1996	Encuesta Permanente de hogares de Propositos Multiples	Direccion General de Estadistica y Censos	9471	National
Czech	1992	Microcensus	Czech Statistical Office	16234	National
Denmark	1992	Income Tax Survey	National Institute of Social Research	12895	National
Ecuador	1996	Encuesta Periodica de Empleo y Desempleo en el Area Urbana	Instituto nacional de Estadistica y Censos	8153	Urban
Finland	1995	Income Distribution Survey	Statistics Finland	9262	National
France	1994	Enquête Budget des familles	INSEE Division Conditions de vie des Ménages	11294	National
Germany	1994	German Social Economic Panel Study	DIW Berlin	6045	National
Hungary	1994	Hungarian Household Panel	Endre Sik / Istvan Toth	1992	National
Israel	1992	Family Expenditure Survey	Israeli Central Bureau of Statistics	5212	National
Italy	1995	Indagine Campionaria sui Bilanci Delle Famiglie	Ufficio Informazioni Statistiche	8135	National
Luxemb	1994	Liewen zu Letzebuerg	Centre d'Etudes de Populations, de Pauvreté et de Politiques Socio-Economiques	1813	National
Mexico	1996	Encuesta Nacional de Increso Gasto de los Hogares	Instituto Nacional de Estadistica, Geografia e Informatica	14042	National
Nether	1994	Socio-Economic Panel	Centraal Bureau voor de Statistiek	5187	National
Norway	1995	Income and Property Distribution Survey	Statistics Norway	10127	National
Panama	1997	Encuesta de Hogares	Direccion de Estadistica y Censo	9897	National
Paraguay	1998	Encuesta Integrada de Hogares	Direccion General de Estadistica, Encuestas y Censos	4353	National
Peru	1997	Encuesta de Hogares	Instituto Nacional de Estadistica e Informatica	3843	National
Poland	1992	Household Budget Survey	Central Statistical Office	6602	National
Slovakia	1992	Slovak Microcensus	Statistical Office of the Slovak Republic Division of SocialStatistics and Demography	17714	National
Spain	1990	Expenditure and Income Survey	Instituto Nacional de Estadistica	11294	National
Sweden	1995	Inkomstfördelningsundersokningen	Statistics Sweden Program for Income and Wealth	16260	National
Taiwan	1995	Survey of Personal Income Distribution	Academia Sinica	14706	National
Uruguay	1996	Encuesta Continua de Hogares	instituto nacional de Estadistica	19322	Urban
USA	1994	March Current Population Survey	Bureau of Labor Statistics	66014	National
Venezuela	1996	Encuesta de Hogares por Mustreo	Oficina Central de Estadistica e Informatica	16323	National

Table B2: Education Thresholds

Name	Years of Schooling Beyond Which A Person Qualifies as Skilled	The LIS Education Level Beyond Which A Person Qualifies as Skilled
Australia	12	Basic/Skilled Vocational Qualification
Argentina	12	N/A
Belgium	12	2nd Level Upper Professional/Technical/ General; Other 2nd Level Upper
Bolivia	12	N/A
Brazil	11	N/A
Britain	13	A Level
Canada	12	Grade 11-13; High School Grad.
Chile	11	N/A
Columbia	11	N/A
Costa Rica	11	N/A
Czech	12	Secondary General/Professional
Denmark	10	Level 2,2nd Stage
Ecuador	12	N/A
Finland*	12	N/A
France	12	Second Stage of Secondary
Germany	10	Secondary
Hungary	12	Secondary
Israel*	12	N/A
Italy	12	High School
Luxembourg	12	Higher Secondary Education
Mexico	11	N/A
Nether	12	Secondary Higher
Norway*	12	N/A
Panama	12	N/A
Paraguay	12	N/A
Peru	11	N/A
Poland	12	Complete Secondary
Slovakia	12	Secondary/Secondary Special/Skilled with Leaving Exam
Spain	12	Secondary Education/Basic Tech. Edu.
Sweden	12	Secondary School
Taiwan	10,12	Senior High/Vocational Graduate
Uruguay	11	N/A
USA	12	High School Diploma
Venezuela	11	N/A

* Finland, Israel and Norway report years of education